

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

OM nucleic - nucleic search, using sw model

Run on: September 2, 2004, 00:17:05 ; Search time 7415 Seconds  
(without alignments)  
11024.270 Million cell updates/sec

Title: US-09-423-575-1  
Perfect score: 1886  
Sequence: 1 attagtataaaattgtgc.....ttaaaaaaaaaaaaaaaaaa 1886

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl : \*

1: gb\_ba : \*  
2: gb\_htg : \*  
3: gb\_in : \*  
4: gb\_om : \*  
5: gb\_ov : \*  
6: gb\_pat : \*  
7: gb\_ph : \*  
8: gb\_pl : \*  
9: gb\_pr : \*  
10: gb\_ro : \*  
11: gb\_scs : \*  
12: gb\_sy : \*  
13: gb\_un : \*  
14: gb\_vi : \*  
15: em\_ba : \*  
16: em\_fun : \*  
17: em\_hum : \*  
18: em\_in : \*  
19: em\_mu : \*  
20: em\_om : \*  
21: em\_or : \*  
22: em\_ov : \*  
23: em\_pat : \*  
24: em\_ph : \*  
25: em\_pl : \*  
26: em\_ro : \*  
27: em\_scs : \*  
28: em\_un : \*  
29: em\_vi : \*  
30: em\_vig\_hum : \*  
31: em\_hcg\_inv : \*  
32: em\_hcg\_other : \*  
33: em\_hcg\_mus : \*  
34: em\_hcg\_pln : \*  
35: em\_hcg\_rod : \*  
36: em\_hcg\_mam : \*  
37: em\_hcg\_vrt : \*  
38: em\_sy : \*  
39: em\_hgo\_hum : \*  
40: em\_hgo\_mus : \*  
41: em\_hgo\_other : \*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Match	Length	DB	ID	Description
1	1886	100.0	1886	8	ATH1RNA	X80126 A.thaliana
2	1882.8	99.8	1886	6	AX002854	AX002854 Sequence
3	1846	97.9	2277	8	AY072174	AY072174 Arabidops
4	1476	78.3	9359	6	AX384394	AX384394 Sequence
5	1476	78.3	9359	6	AX473364	AX473364 Sequence
6	1453	77.0	1453	8	AY098513	AY098513 Arabidops
7	1110.6	58.9	4878	8	ATH1GEN	X80127 A.thaliana
c 8	1110.6	58.9	94482	8	ATF26221	AL031804 Arabidops
c 9	1110.6	58.9	198669	8	ATCHRIV78	AL161582 Arabidops
10	702.8	37.3	10072	8	ATZ83320	Z83320 A.thaliana
11	255	13.5	1182	8	AF375967	AF375967 Lycopersi
12	132	7.0	2065	8	AF406703	AF406703 Solanum t
13	128.6	6.8	2735	8	AF406697	AF406697 Solanum t
14	127	6.7	2636	8	AF375966	AF375966 Lycopersi
15	126.4	6.7	2570	8	AK067939	AK067939 Oryza sat
16	126.4	6.7	2630	8	AK069994	AK069994 Oryza sat
17	126.2	6.7	1898	8	AF406698	AF406698 Solanum t
18	125.6	6.7	2128	8	AF406702	AF406702 Solanum t
19	124.2	6.6	2143	8	AK109849	AK109849 Oryza sat
20	124	6.6	1089	8	AK061393	AK061393 Oryza sat
21	124	6.6	1089	8	AK104713	AK104713 Oryza sat
22	123.6	6.6	2043	6	AX506642	AX506642 Sequence
23	123.6	6.6	2043	8	BT008422	BT008422 Arabidops
24	123.6	6.6	2074	8	AY096548	AY096548 Arabidops
25	123.6	6.6	2385	8	AF353094	AF353094 Arabidops
26	123.6	6.6	2572	8	AY072175	AY072175 Arabidops
27	123.6	6.6	2578	8	AY099854	AY099854 Arabidops
28	122.6	6.5	1731	8	AF406700	AF406700 Solanum t
29	121.8	6.5	1599	6	AX507459	AX507459 Sequence
30	121.4	6.4	2059	8	AY052366	AY052366 Arabidops
31	119.8	6.4	1443	6	AX412914	AX412914 Sequence
32	119.8	6.4	1443	8	BT000609	BT000609 Arabidops
33	119.8	6.4	2126	8	AY056796	AY056796 Arabidops
34	117.8	6.2	1907	8	AF375964	AF375964 Lycopersi
35	115	6.1	1896	8	AK104872	AK104872 Oryza sat
36	115	6.1	2476	8	AK121356	AK121356 Oryza sat
37	115	6.1	3030	8	AK070465	AK070465 Oryza sat
38	114.8	6.1	1939	8	AF406701	AF406701 Solanum t
39	114.4	6.1	1884	6	AX506210	AX506210 Sequence
40	114.4	6.1	1884	8	AY142019	AY142019 Arabidops
41	114.4	6.1	2050	8	AF353092	AF353092 Arabidops
42	114.4	6.1	2124	8	AY058179	AY058179 Arabidops
43	114	6.0	2952	8	AK070543	AK070543 Oryza sat
44	113.4	6.0	2545	8	AF173816	AF173816 Arabidops
45	113.4	6.0	3162	8	AY050459	AY050459 Arabidops

ALIGNMENTS

RESULT 1

ATH1RNA

LOCUS: ATH1RNA

DEFINITION: A.thaliana Ath1 mRNA.

ACCESSION: X80126

VERSION: X80126.1

KEYWORDS: H1 gene.

SOURCE: Arabidopsis thaliana (thale cress)

ORGANISM: Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE 1

AUTHORS: Quaedyveling, N., Dockx, J., Rook, F., Weisbeek, P. and Smeekens, S.

TITLE: The homeobox gene ATH1 of Arabidopsis is derepressed in the

1886 bp mRNA linear PLN 02-MAR-1995



Db	1501	CCGAAAGATTGGAGAAACAATCTTCTAGCTATTACGAAGTGCGTTGACAAGAAGTCAGGTA	1560
Qy	1561	TCAACTGCGTTTATAAATGCGCGGTTAGCTATGGAAGCCGATGATAGAAAGAGATGTAT	1620
Db	1561	TCAACTGCGTTTATAAATGCGCGGTTAGCTATGGAAGCCGATGATAGAAAGAGATGTAT	1620
Qy	1621	GCGGAAATGAACAGAGGAAAGCTCAATAACAGTCACATTCACCCCAACGGACCAATCTTT	1680
Db	1621	GCGGAAATGAACAGAGGAAAGCTCAATAACAGTCACATTCACCCCAACGGACCAATCTTT	1680
Qy	1681	CGAATGCCAAAATCTGTTATGATGAGCCCAAGCAATGCATAAATAAGACAACAATTGTGTT	1740
Db	1681	CGAATGCCAAAATCTGTTATGATGAGCCCAAGCAATGCATAAATAAGACAACAATTGTGTT	1740
Qy	1741	TAGCAACTTTTGTCATAATTAAGCAAAATGCTACTCTATGATGCCCCAAAACCTAAACCATG	1800
Db	1741	TAGCAACTTTTGTCATAATTAAGCAAAATGCTACTCTATGATGCCCCAAAACCTAAACCATG	1800
Qy	1801	TAGACATCATTAACGTATGCTTAATTTGTAATATACAACTCCCTTTATCTTTGACTATTTTC	1860
Db	1801	TAGACATCATTAACGTATGCTTAATTTGTAATATACAACTCCCTTTATCTTTGACTATTTTC	1860
Qy	1861	ATTTTATTAATAAAAAAAAAAAAAAAAAAAAAA	1896
Db	1861	ATTTTATTAATAAAAAAAAAAAAAAAAAAAAAA	1896

RESULT 2			
AX002854			
LOCUS	AX002854	1886 bp	DNA
DEFINITION	Sequence 1 from Patent WO9851800.	linear	PAT 07-SEP-2000

ACCESSION AX002854  
 VERSION AX002854.1 GI:9885172  
 KEYWORDS  
 SOURCE Arabidopsis thaliana (thale cress)

SOURCE  
Arabidopsis thaliana (thale cress)

ORGANISM *Arabidopsis thaliana*

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:

rosids: eucosids II: Brassicales: Brassicaceae: Arabidopsis.

## REFERENCE

PROVENIERS, M., Smeekens, S. and Weisbeek, P.

**TITLE**  
Plant gene constructs and their use

**JOURNAL**  
**Patent: WO 9851800-A 1 19-NOV-1998:**

COÖRDMINE  
TACHTS, NO 2021000-A 1 19-NOV-1998,  
HAVE D J VAN DER BV (NL): PROVENTERS MARCEL. (NL): SMEEKENS STEF

LEAVE D O VAN DER BV (NL); FROVENTERS MARCEL (NL); SMEENENS SUUF (NL); WEITSBEEK PETER (NL.)

FEATURES	Location/Qualifiers	(NE); WEISBERG FETTER (NJ)
1. The first feature is the name of the person who is the subject of the document.		
2. The second feature is the name of the person who is the author of the document.		
3. The third feature is the name of the person who is the editor of the document.		
4. The fourth feature is the name of the person who is the publisher of the document.		
5. The fifth feature is the name of the person who is the distributor of the document.		
6. The sixth feature is the name of the person who is the reviewer of the document.		
7. The seventh feature is the name of the person who is the critic of the document.		
8. The eighth feature is the name of the person who is the commentator on the document.		
9. The ninth feature is the name of the person who is the analyst of the document.		
10. The tenth feature is the name of the person who is the expert on the document.		
11. The eleventh feature is the name of the person who is the specialist in the document.		
12. The twelfth feature is the name of the person who is the authority on the document.		
13. The thirteenth feature is the name of the person who is the leader in the document.		
14. The fourteenth feature is the name of the person who is the pioneer in the document.		
15. The fifteenth feature is the name of the person who is the innovator in the document.		
16. The sixteenth feature is the name of the person who is the creator in the document.		
17. The seventeenth feature is the name of the person who is the originator in the document.		
18. The eighteenth feature is the name of the person who is the inventor in the document.		
19. The nineteenth feature is the name of the person who is the discoverer in the document.		
20. The twentieth feature is the name of the person who is the finder in the document.		
21. The twenty-first feature is the name of the person who is the seeker in the document.		
22. The twenty-second feature is the name of the person who is the explorer in the document.		
23. The twenty-third feature is the name of the person who is the adventurer in the document.		
24. The twenty-fourth feature is the name of the person who is the risk-taker in the document.		
25. The twenty-fifth feature is the name of the person who is the gambler in the document.		
26. The twenty-sixth feature is the name of the person who is the speculator in the document.		
27. The twenty-seventh feature is the name of the person who is the investor in the document.		
28. The twenty-eighth feature is the name of the person who is the entrepreneur in the document.		
29. The twenty-ninth feature is the name of the person who is the businessperson in the document.		
30. The thirtieth feature is the name of the person who is the professional in the document.		
31. The thirty-first feature is the name of the person who is the expert in the document.		
32. The thirty-second feature is the name of the person who is the specialist in the document.		
33. The thirty-third feature is the name of the person who is the authority in the document.		
34. The thirty-fourth feature is the name of the person who is the leader in the document.		
35. The thirty-fifth feature is the name of the person who is the pioneer in the document.		
36. The thirty-sixth feature is the name of the person who is the innovator in the document.		
37. The thirty-seventh feature is the name of the person who is the creator in the document.		
38. The thirty-eighth feature is the name of the person who is the originator in the document.		
39. The thirty-ninth feature is the name of the person who is the inventor in the document.		
40. The fortieth feature is the name of the person who is the discoverer in the document.		
41. The forty-first feature is the name of the person who is the finder in the document.		
42. The forty-second feature is the name of the person who is the seeker in the document.		
43. The forty-third feature is the name of the person who is the explorer in the document.		
44. The forty-fourth feature is the name of the person who is the adventurer in the document.		
45. The forty-fifth feature is the name of the person who is the risk-taker in the document.		
46. The forty-sixth feature is the name of the person who is the gambler in the document.		
47. The forty-seventh feature is the name of the person who is the speculator in the document.		
48. The forty-eighth feature is the name of the person who is the investor in the document.		
49. The forty-ninth feature is the name of the person who is the entrepreneur in the document.		
50. The fiftieth feature is the name of the person who is the businessperson in the document.		
51. The fifty-first feature is the name of the person who is the professional in the document.		
52. The fifty-second feature is the name of the person who is the expert in the document.		
53. The fifty-third feature is the name of the person who is the specialist in the document.		
54. The fifty-fourth feature is the name of the person who is the authority in the document.		
55. The fifty-fifth feature is the name of the person who is the leader in the document.		
56. The fifty-sixth feature is the name of the person who is the pioneer in the document.		
57. The fifty-seventh feature is the name of the person who is the innovator in the document.		
58. The fifty-eighth feature is the name of the person who is the creator in the document.		
59. The fifty-ninth feature is the name of the person who is the originator in the document.		
60. The sixtieth feature is the name of the person who is the inventor in the document.		
61. The sixty-first feature is the name of the person who is the discoverer in the document.		
62. The sixty-second feature is the name of the person who is the finder in the document.		
63. The sixty-third feature is the name of the person who is the seeker in the document.		
64. The sixty-fourth feature is the name of the person who is the explorer in the document.		
65. The sixty-fifth feature is the name of the person who is the adventurer in the document.		
66. The sixty-sixth feature is the name of the person who is the risk-taker in the document.		
67. The sixty-seventh feature is the name of the person who is the gambler in the document.		
68. The sixty-eighth feature is the name of the person who is the speculator in the document.		
69. The sixty-ninth feature is the name of the person who is the investor in the document.		
70. The seventieth feature is the name of the person who is the entrepreneur in the document.		
71. The seventy-first feature is the name of the person who is the businessperson in the document.		
72. The seventy-second feature is the name of the person who is the professional in the document.		
73. The seventy-third feature is the name of the person who is the expert in the document.		
74. The seventy-fourth feature is the name of the person who is the specialist in the document.		
75. The seventy-fifth feature is the name of the person who is the authority in the document.		
76. The seventy-sixth feature is the name of the person who is the leader in the document.		
77. The seventy-seventh feature is the name of the person who is the pioneer in the document.		
78. The seventy-eighth feature is the name of the person who is the innovator in the document.		
79. The seventy-ninth feature is the name of the person who is the creator in the document.		
80. The eightieth feature is the name of the person who is the originator in the document.		
81. The eighty-first feature is the name of the person who is the inventor in the document.		
82. The eighty-second feature is the name of the person who is the discoverer in the document.		
83. The eighty-third feature is the name of the person who is the finder in the document.		
84. The eighty-fourth feature is the name of the person who is the seeker in the document.		
85. The eighty-fifth feature is the name of the person who is the explorer in the document.		
86. The eighty-sixth feature is the name of the person who is the adventurer in the document.		
87. The eighty-seventh feature is the name of the person who is the risk-taker in the document.		
88. The eighty-eighth feature is the name of the person who is the gambler in the document.		
89. The eighty-ninth feature is the name of the person who is the speculator in the document.		
90. The ninetieth feature is the name of the person who is the investor in the document.		
91. The ninety-first feature is the name of the person who is the entrepreneur in the document.		
92. The ninety-second feature is the name of the person who is the businessperson in the document.		
93. The ninety-third feature is the name of the person who is the		

FEATURES	LOCATION/QUALITIES
SOURCE	1 1886

237095  
T. .1886  
/organism="Arabidopsis thaliana"

```
/organism="Arabidopsis thaliana"  
/mol_type="unassigned DNA"
```

```

/mot_type="unassigned DNA"
/ab_xref="Favon. 2702"

```

```
OPTGTTN
/ab_xref="taxon:3702"
```

# ORIGIN

[illegible]

Query Match	Score	DB	Length
Best local similarity	99.88	1882.8	6
Best local similarity	99.88	1886	6

Best Local Similarity 99.9%; pred: No. 0;  
Matches 1994: Concentration 0.0; Mismatched 0.0; Total 1.0

```
Matches 1884; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

1 ATTTAGTTATAAAATGTTGCTATTTTGTGATCTAGTGCTCTGAACTTTTAGTGAGCAG 60  
QY

100

DD 1 ATTTAGTTATAAAATGTTGCTATTTGTTGATCTAGTCTCTGAATCTTTTAGTGAGGCAG 60

**Table 1**

61 ATGATGAAGATTATGAAATTCCTTCATGAAATTATTGTAAGAAAAGAACATAGAGAAGCT 120

100

D5 61 ATGATGAAGATTATGAAATTCCTCATGAAATTATTGTAAGAAAGAACATAGAGAAGCT 120

QY 121 GCGAATGAAGTACACTGTTCTTCACGAGAAAGATAAATAAGCATTATCTTCTT 180

100

D5 121 GCGAATGAAAGTACACTGTTCTTCACGGAGAAAGATAAATAAGCATTATCTTCTT 180

QY 181 CTTCAGTTTTTAACACACACATTTTGGAAATTTTGATGTAAATACTCTTTGGAACGTTGT 240

100

D<sub>b</sub> 181 CTTCAGTTTTAACACACACATTTTGGAAATTTTGATGTAAAAATCTCTTTGGAAACGTTGT 240

Qy 241 GTTGTCTGAAATCTTCCCAAGGTTCTATCAGAAGAAGGATAAAGTTTCATAGAAC 300

D	b	241	GTTGTCTGAAATCTTCCAAAGTGTTCTATCAGAAGAGAGGATAAAGTTTCATGAACAAC	300
Q	y	301	CCAATGGACAACAACAACAACAACACTTTTAGTTTCTCGATTAATGTCATGACTAAC	360
D	b	301	CCAATGGACAACAACAACAACAACACTTTTAGTTTCTCGATTAATGTCATGACTAAC	360
Q	y	361	CAAAATCCTCTTCTCATGGATTATACCTTCAAGAGAAGATTCAAATTCATTTCAACA	420
D	b	361	CAAAATCCTCTTCTCATGGATTATACCTTCAAGAGAAGATTCAAATTCATTTCAACA	420
Q	y	421	ATGCTTCCATGGAAACCACATCAGATCAGATCCCTCAAAATGGTGGCTTGATATTTTC	480
D	b	421	ATGCTTCCATGGAAACCACATCAGATCAGATCCCTCAAAATGGTGGCTTGATATTTTC	480
Q	y	481	AATTTATGCTGACTAACCAATATCTTATCATCTTCTCCAGGPTCTATCGATGTTCAAGAT	540
D	b	481	AATTTATGCTGACTAACCAATATCTTATCATCTTCTCCAGGPTCTATCGATGTTCAAGAT	540
Q	y	541	AACCGCAATGTTGAGTTTCATGGTCTCTCTCTCATCTCTCTCCACTTCATCTTTGGAT	600
D	b	541	AACCGCAATGTTGAGTTTCATGGTCTCTCTCTCTCATCTCTCTCCACTTCATCTTTGGAT	600
Q	y	601	CATTTAAGACATATGATGATTCCTCAAAACAATGTGGGTTTTTCAAGCAAAATAGTGAG	660
D	b	601	CATTTAAGACATATGATGATTCCTCAAAACAATGTGGGTTTTTCAAGCAAAATAGTGAG	660
Q	y	661	TTTCAGGCATTTTCAGGTGATGTTGGTCCAAAGTGAACCAATGATGTCATTCGGTGAA	720
D	b	661	TTTCAGGCATTTTCAGGTGATGTTGGTCCAAAGTGAACCAATGATGTCATTCGGTGAA	720
Q	y	721	GAAGATTTCCGGTTTTCTAATTTCCGAATAAAGAAAACAATGAGCTTTTCATTGAGCTTGCA	780
D	b	721	GAAGATTTCCGGTTTTCTAATTTCCGAATAAAGAAAACAATGAGCTTTTCATTGAGCTTGCA	780
Q	y	781	TCAGATGTTTCTGATGAATGCTCGGAGATAAGTCTTTTGTGACGCTACAAGATTAGCCTCA	840
D	b	781	TCAGATGTTTCTGATGAATGCTCGGAGATAAGTCTTTTGTGACGCTACAAGATTAGCCTCA	840
Q	y	841	GAGCAAGCTTCTGACGACGAAAGACATTTCTTAATAACGTTGTGTTACTCAAGGTTTCTCT	900
D	b	841	GAGCAAGCTTCTGACGACGAAAGACATTTCTTAATAACGTTGTGTTACTCAAGGTTTCTCT	900
Q	y	901	CAACTTATATTTGGCTCAAAATACCTTCACTCTGTTTCAAGAAATACTATCTCATTTCCGCC	960
D	b	901	CAACTTATATTTGGCTCAAAATACCTTCACTCTGTTTCAAGAAATACTATCTCATTTCCGCC	960
Q	y	961	GCATATCGCTCGATTTATCATCTCGAGAAACCGAGTCAGGAGCTGCTAGTTTCAGCCTTT	1020
D	b	961	GCATATCGCTCGATTTATCATCTCGAGAAACCGAGTCAGGAGCTGCTAGTTTCAGCCTTT	1020
Q	y	1021	ACTTTCAGTTTTGAGAAATATACTGAGTTTCTTGATGGTGATTCTAATAACTCGGAGGCG	1080
D	b	1021	ACTTTCAGTTTTGAGAAATATACTGAGTTTCTTGATGGTGATTCTAATAACTCGGAGGCG	1080
Q	y	1081	GGTTTCGGATCTACATTTCAAAGGAGAGCATTTAGAAGCAAAAGAAAACCCATCTCTTTGGAT	1140
D	b	1081	GGTTTCGGATCTACATTTCAAAGGAGAGCATTTAGAAGCAAAAGAAAACCCATCTCTTTGGAT	1140
Q	y	1141	CTTCTTCAATGGTGGATGATCGATATAGTTCATTGCGTAGATGAGATTCATACGGTTATA	1200
D	b	1141	CTTCTTCAATGGTGGATGATCGATATAGTTCATTGCGTAGATGAGATTCATACGGTTATA	1200
Q	y	1201	TCAGCGTTCATGCTGCAACCGAGTTTAGATCCACAGTTTACACACCGCGTTTCCCTCCAA	1260
D	b	1201	TCAGCGTTCATGCTGCAACCGAGTTTAGATCCACAGTTTACACACCGCGTTTCCCTCCAA	1260
Q	y	1261	ACCGTTTCTCTTTATACAGAACCTTGAGAGAGAGAACTCTGAAGAGAGATAATCTCTATG	1320
D	b	1261	ACCGTTTCTCTTTATACAGAACCTTGAGAGAGAGAACTCTGAAGAGAGATAATCTCTATG	1320
Q	y	1321	GGATCTGTATTTGAGAGAGGCGAAAGACAAGACTCAAGAAACCTCTATGTTCCACCAAGCAT	1380

1321 GGATCTGTATTGGAGAGAGGCAAGCAAGACTCAAGAAACCTCTATTTCCACAGCAT 1380  
Db  
1381 TGCCTTCTTCAGCAGCTGAACCAAGAAAGAACCATCATGATTTGGAGAGCTCAACGAGGTTG 1440  
QY  
1381 TGCCTTCTTCAGCAGCTGAACCAAGAAAGAACCATCATGATTTGGAGAGCTCAACGAGGTTG 1440  
Db  
1441 CTGAGAAATCTGTTTCGGTTCTACGGAATTCGGATTCGAAATTCCTTCCACCCITAC 1500  
QY  
1441 CTGAGAAATCTGTTTCGGTTCTACGGAATTCGGATTCGAAATTCCTTCCACCCITAC 1500  
Db  
1501 CGGAAAGATTCGGAGAAACATCTTCTAGCTATACCAAGTGGCTTACCAAGAAAGTCAGGTA 1560  
QY  
1501 CGGAAAGATTCGGAGAAACATCTTCTAGCTATACCAAGTGGCTTACCAAGAAAGTCAGGTA 1560  
Db  
1561 TCAAACTGTTTATAAATCGCGGTTAGGCTATGGAAGCCGATGATAGAGAGATGAT 1620  
QY  
1561 TCAAACTGTTTATAAATCGCGGTTAGGCTATGGAAGCCGATGATAGAGAGATGAT 1620  
Db  
1621 GCGGAATCAACAGGAGAGCTCAATACAGTCACTTCAACCAAGCCGACCACTCTT 1680  
QY  
1621 GCGGAATCAACAGGAGAGCTCAATACAGTCACTTCAACCAAGCCGACCACTCTT 1680  
Db  
1681 CGAATGCCAAAATCTGTTATGATGAGCAAGCAATGCATATAAGACAACTTGT 1740  
QY  
1681 CGAATGCCAAAATCTGTTATGATGAGCAAGCAATGCATATAAGACAACTTGT 1740  
Db  
1741 TACCAACTTTGATTAATAGCAATGCTCTCTATGATTCGCCAAACCTTAAACCATG 1800  
QY  
1741 TACCAACTTTGATTAATAGCAATGCTCTCTATGATTCGCCAAACCTTAAACCATG 1800  
Db  
1801 TAGCACTATCATTCAGTATGTTATATGATATATACCACTCTTATCTTTCGACTATTC 1860  
QY  
1801 TAGCACTATCATTCAGTATGTTATATGATATATACCACTCTTATCTTTCGACTATTC 1860  
Db  
1861 ATTTTATTAATAAAAAAAAAAAAAA 1886  
QY  
1861 ATTTTATTAATAAAAAAAAAAAAAA 1886  
Db

RESULT 3  
AY072174  
LOCUS  
DEFINITION  
Arabidopsis thaliana putative homeobox gene ATH1 protein  
(At4g32980) mRNA, complete cds.  
ACCESSION  
AY072174  
VERSION  
AY072174.1  
KEYWORDS  
FLI CDNA.  
SOURCE  
Arabidopsis thaliana (thale cress)  
ORGANISM  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 2277)  
Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,  
Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,  
Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,  
Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,  
Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,  
Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,  
Davis,R.W., Ecker,J.R. and Theologis,A.  
Arabidopsis Full Length cDNA Clones  
Unpublished  
2 (bases 1 to 2277)  
Yamada,K., Banh,J., Chan,M.M., Chang,C.H., Chang,E., Dale,J.M.,  
Deng,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L.,  
Tang,C.C., Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L.,  
Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,  
Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,  
Lin,J., Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,  
Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A.,  
Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.  
Direct Submission  
Submitted (02-JAN-2002) Plant Gene Expression Center, 800 Buchanan

Street, Albany, CA 94710, USA  
RIKEN Genomic Sciences Center (GSC) members carried out the  
collection and clustering of RAFL cDNAs (RAFL cDNA : RIKEN  
Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J.,  
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,  
Hayashizaki,Y. and Shinozaki,K.  
The Salk, Stanford, PCEC (SSP) Consortium members carried out the  
sequencing and annotation of the RAFL cDNAs: Yamada,K., Banh,J.,  
Chan,M.M., Chang,C.H., Chang,E., Dale,J.M., Deng,J.M.,  
Goldsmith,A.D., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C.,  
Toriumi,M., Wu,H.C., Yamamura,Y., Yu,G., Bowser,L.,  
Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Lam,B., Lin,J.,  
Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J., Shinn,P.,  
Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.  
Yamada,K. (SSP/PGSC) and Seki,M. (RIKEN GSC) contributed equally to  
this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)  
contributed equally to this work as PIs.  
Annotation is based on the January 2002 version of the Arabidopsis  
genome submitted to GenBank.  
FEATURES  
Location/Qualifiers  
source  
1..2277  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/chromosome="4"  
/clone="RAFL09-24-021 (R18486)"  
/ecotype="Columbia"  
/notes="This clone is in a modified pBluescript vector  
(FLC-1) as a BamHI/XhoI insert."  
gene  
1..2277  
/gene="At4g32980"  
5'UTR  
1..707  
/gene="At4g32980"  
misc\_difference 2..3  
/gene="At4g32980"  
/notes="compared to genomic sequence"  
/replace="gta"  
misc\_difference 373  
/gene="At4g32980"  
/notes="not present in genomic sequence"  
misc\_difference 595  
/gene="At4g32980"  
/notes="not present in genomic sequence"  
CDS  
708..2129  
/gene="At4g32980"  
/codon\_start=1  
/evidence=experimental  
/product="putative homeobox gene ATH1 protein"  
/protein\_id="AAU5996.1"  
/db\_xref="GI:18176168"  
translation="MDNNNNNTFSSLDNVNTNQNPFLMDFPSREDSTSFSTMLPWN  
TIRSDLOMGDFIDFNSMLNTKYLSSPSIDVQNNRVFMAPPHPPLPLHLR  
HYDSSNNWGPENSEFOAFSGVGPSPMSTGEEDFPELISKNRNEILSLAS  
DVSECEISLCAATLASEQSCSKDISNNVTGQFSQSLIGSKYLHSVQELISHP  
AAYSLDSRGTSAGAASTAFENITFELDGSSNNEAGFGSTFQRALEAKTH  
LDLQGVDRYSHCVDETHVISAFAATELDPQLHTFALQTVFLYKNJRKICK  
KIISGVLERGDKQETSMFQHCLLQQLKRNQHWIRPQRGPEKSVSVLRNMF  
QNFHPYKDSKHLAIRSGLRISQVSNFVARNVRLWPKMIEEMYNKRLNLS  
HIQNGPTLRMPKSVNMSQAMHK"  
2130..2277  
/gene="At4g32980"  
misc\_difference 2261  
/gene="At4g32980"  
/notes="not present in genomic sequence"  
ORIGIN  
Query Match 97.9%; Score 1846; DB 8; Length 2277;  
Best Local Similarity 99.4%; Pred No. 0;  
Matches 1863; Conservative 0; Mismatches 10; Indels 1; Gaps 1;  
Direct Submission



QY 1 ATTAGTATAAATGTTGCTATTTGTTGATCTAGTCTGAACTTTTGTAGTGAGCAG 60  
DB 404 ATTAGTATAAATGTTGCTATTTGTTGATCTAGTCTGAACTTTTGTAGTGAGCAG 463  
QY 61 ATGATGAAGATTATGAATTTCTCATGAATTTATTTGTAAGAAAAGAAATAGAGAGCT 120  
DB 464 ATGATGAAGATTATGAATTTCTCATGAATTTATTTGTAAGAAAAGAAATAGAGAGCT 523  
QY 121 GCGGAATGAAGTACACTGTTCTTTCCAGGAGAAAAGATAAATAAGCAATTATCTCTT 180  
DB 524 GCGGAATGAAGTACACTGTTCTTTCCAGGAGAAAAGATAAATAAGCAATTATCTCTT 583  
QY 181 CTTTCAG-TTTTTAACAACATTTTGGAAATTTTGAATTTTGAATTTTCTTTGGAAAGTTG 239  
DB 584 CTTTCAGTTTTTAAACAACATTTTGGAAATTTTGAATTTTGAATTTTCTTTGGAAAGTTG 643  
QY 240 TGTGTTCTGAAATCTTTCCCAAGGTTCTATCAAGAAAGAAAGATAAAGTTTTCATAGAAA 299  
DB 644 TGTGTTCTGAAATCTTTCCCAAGGTTCTATCAAGAAAGAAAGATAAAGTTTTCATAGAAA 703  
QY 300 CCAATATGGACAAACAAACAAACAACTTTTGTCTCTGATAAATGTATGACTAA 359  
DB 704 CCAATATGGACAAACAAACAAACAACTTTTGTCTCTGATAAATGTATGACTAA 763  
QY 360 CCAAAATCTCTTCTCATGATTTTATACCTTCAAGAGAGATTCAACTTCATCTCAAC 419  
DB 764 CCAAAATCTCTTCTCATGATTTTATACCTTCAAGAGAGATTCAACTTCATCTCAAC 823  
QY 420 AATGCTTCCATGAATACCATCAGATCAGATCTCTACAAATGGGTGGCTTTGATATTTT 479  
DB 824 AATGCTTCCATGAATACCATCAGATCAGATCTCTACAAATGGGTGGCTTTGATATTTT 883  
QY 480 CAATCTATGCTGACTAAACAATACCTTATCACTTCTCCACAGGTCTATCGATGTTCAAGA 539  
DB 884 CAATCTATGCTGACTAAACAATACCTTATCACTTCTCCACAGGTCTATCGATGTTCAAGA 943  
QY 540 TAAACCAATGTTGAGTTTATGCTCTCTCTCATCTCTCTCCACATTCATCTCTTGA 599  
DB 944 TAAACCAATGTTGAGTTTATGCTCTCTCTCATCTCTCTCCACATTCATCTCTTGA 1003  
QY 600 TCATTTAAGACATATGATGATTCCTCAACAAACATGTTGGGTTTGAAGCAAAATAGTGA 659  
DB 1004 TCATTTAAGACATATGATGATTCCTCAACAAACATGTTGGGTTTGAAGCAAAATAGTGA 1063  
QY 660 GTTTCAGGCAATTTTCAGGTGATGTTGGTCCAAAGTGAACCAATGATCTACATTCGGTGA 719  
DB 1064 GTTTCAGGCAATTTTCAGGTGATGTTGGTCCAAAGTGAACCAATGATCTACATTCGGTGA 1123  
QY 720 AGAAGATTTCCGTTTCTAATTTTGAATAAAGAAACAATGAGCTTTTCAATGAGTCTTGC 779  
DB 1124 AGAAGATTTCCGTTTCTAATTTTGAATAAAGAAACAATGAGCTTTTCAATGAGTCTTGC 1183  
QY 780 ATCAGATGTTTCTGATGATGCTCGAGATAGTCTTGTGAGCTACAGATTAAGCTC 839  
DB 1184 ATCAGATGTTTCTGATGATGCTCGAGATAGTCTTGTGAGCTACAGATTAAGCTC 1243  
QY 840 AGAGCAAGCTTCTTGACAGCAAAAGACATTTCTAATAACGTTGTTTACTCAAGGTTTCTC 899  
DB 1244 AGAGCAAGCTTCTTGACAGCAAAAGACATTTCTAATAACGTTGTTTACTCAAGGTTTCTC 1303  
QY 900 TCAACTTATATTGGCTCAAAATACCTTCACTCTGTTTCAAGAAAATACATCTATCTATTCGC 959  
DB 1304 TCAACTTATATTGGCTCAAAATACCTTCACTCTGTTTCAAGAAAATACATCTATCTATTCGC 1363  
QY 960 CGCATACTCTCGATTTATCTCTCGAGGAACCGAGTCAGAGCTGCTAGTTCAGGCTT 1019  
DB 1364 CGCATACTCTCGATTTATCTCTCGAGGAACCGAGTCAGAGCTGCTAGTTCAGGCTT 1423  
QY 1020 TACTTCAGCTTTTGAATATAACTGAGTTTCTTGTATGTTGATTTCTTAATACTCGAGGC 1079  
DB 1424 TACTTCAGCTTTTGAATATAACTGAGTTTCTTGTATGTTGATTTCTTAATACTCGAGGC 1483  
QY 1080 GGGTTTCGGATCTACATTTCAAGAGGAGCATTTAGAGCAAGAAAACCCATCTCTTGA 1139

DB 1484 GGGTTTCGGATCTACATTTCAAGAGGAGCATTTAGAGCAAGAAAACCCATCTCTTGA 1543  
QY 1140 TCTTCTTCAATGTTGGATGATGATAGTATAGTCAATTCGTTAGATGAGATTCATAGGTTAT 1199  
DB 1544 TCTTCTTCAATGTTGGATGATGATAGTATAGTCAATTCGTTAGATGAGATTCATAGGTTAT 1603  
QY 1200 ATCAGCGTTCCATGCTGCAACCGAGTTAGATCCACAGTTACACACCGGTTTGGCCCTCCA 1259  
DB 1604 ATCAGCGTTCCATGCTGCAACCGAGTTAGATCCACAGTTACACACCGGTTTGGCCCTCCA 1663  
QY 1260 AACCGTTTCTCTTATACAAAGAACCTTGAGAGAGAGATCTGCAAGAGAGATAATCTCTAT 1319  
DB 1664 AACCGTTTCTCTTATACAAAGAACCTTGAGAGAGAGATCTGCAAGAGAGATAATCTCTAT 1723  
QY 1320 GGGATCTGTTATGGAGAGAGGCAAGACACTCAAGAAACCTCTATGTTTCCACACAGCA 1379  
DB 1724 GGGATCTGTTATGGAGAGAGGCAAGACACTCAAGAAACCTCTATGTTTCCACACAGCA 1793  
QY 1380 TTGCTCTTTCAGCAGCTGAAACGAAAGAACCAATCAGATTTGGAGAGCTCAACAGAGTTT 1439  
DB 1784 TTGCTCTTTCAGCAGCTGAAACGAAAGAACCAATCAGATTTGGAGAGCTCAACAGAGTTT 1843  
QY 1440 GCCTGAGAAATCTGTTTGGGTTCTACGGAATTTGATGTTTCCAAACCTTCTTCAACCTTA 1499  
DB 1844 GCCTGAGAAATCTGTTTGGGTTCTACGGAATTTGATGTTTCCAAACCTTCTTCAACCTTA 1903  
QY 1500 CCGAAAGATTTCGAGAGAAACATCTTCTAGCTATACGAAGTGGCTTGAACAAGTCAAGT 1559  
DB 1904 CCGAAAGATTTCGAGAGAAACATCTTCTAGCTATACGAAGTGGCTTGAACAAGTCAAGT 1963  
QY 1560 ATCAAACTGTTTATAAATGCGGGTTAGGCTATGNAAGCGGATGATAGAGAGATGTA 1619  
DB 1964 ATCAAACTGTTTATAAATGCGGGTTAGGCTATGNAAGCGGATGATAGAGAGATGTA 2023  
QY 1620 TCGGAAATGAACAGAGAGAGCTCAATAACAGTCAATTCAGCCCAACCGACCAACTCT 1679  
DB 2024 TCGGAAATGAACAGAGAGAGCTCAATAACAGTCAATTCAGCCCAACCGACCAACTCT 2083  
QY 1680 TCGAATGCCAAATCTGTTATGATGAGCCAAAGAAATGCAATAAAGACAAATTTGTTGT 1739  
DB 2084 TCGAATGCCAAATCTGTTATGATGAGCCAAAGAAATGCAATAAAGACAAATTTGTTGT 2143  
QY 1740 TTACCACTTTGTTGATTAATAGGCAATTTGCTACTCTATGTTGCCCAAAACCTTAAACCAT 1799  
DB 2144 TTACCACTTTGTTGATTAATAGGCAATTTGCTACTCTATGTTGCCCAAAACCTTAAACCAT 2203  
QY 1800 GTACGATATCATTAAGTATGTTAATATGTTATATACAACTCTTTTATCTTTGACTATTT 1859  
DB 2204 GTACGATATCATTAAGTATGTTAATATGTTATATACAACTCTCTTTATCTTTGACTAGAA 2263  
QY 1860 CATTTTATTAATA 1873  
DB 2264 AAAAAAAAAAAAAA 2277

## RESULT 4

AX384394  
LOCUS AX384394  
DEFINITION Sequence 3 from Patent WO0214524.  
ACCESSION AX384394  
VERSION AX384394.1 GI:19577678  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1  
AUTHORS van der Valk, P., van Dun, C.M., Smeekens, S.C. and Proveniers, M.C.  
TITLE Inhibition of generative propagation in genetically modified  
herbicide resistant grasses  
JOURNAL Patent: WO 0214524-A 3 21-FEB-2002;  
Advanta Seeds B.V. (NL)  
FEATURES Location/Qualifiers

PAT 19-MAR-2002





Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

# TITLE JOURNAL REFERENCE AUTHORS

Arabidopsis Open Reading Frame (ORF) Clones  
Unpublished  
2 (bases 1 to 1453)  
Yamada, K., Ban, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M., Deng, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, J.M., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

# TITLE JOURNAL COMMENT

Submitted (16-APR-2002) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA  
The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members constructed and sequenced the PUNI (ORF) clones using the RAFL cDNAs: Yamada, K., Ban, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M., Deng, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, J.M., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis genome submitted to GenBank.

# FEATURES source

Location/Qualifiers  
1..1453  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/chromosome="4"  
/clone="U18486"  
/ecotype="Columbia"  
/note="this clone is in PUNI 51."  
1..1453  
/gene="At4g32980"  
1..1422  
/gene="At4g32980"  
/codon\_start=1  
/evidence=experimental  
/product="putative homeobox gene ATH1 protein"  
/protein\_id="AA20163.1"  
/db\_xref="GI:20465955"  
/translation="MDNNNNNTFSSLDNVNTPNPLIMDFIPREDSTGFSTMLPWN  
TIRSDPLQMGDFIDFNSMLTKNLYSSSPSIDVDONRVEFMAPPHPPLHLDHLR  
HYDSSNNMWFSEANSEFCAFVGVVGPSEPMSTFGEDRPFLI SNKRNNLSLSLAS  
DVSDECSILSCATRLASBOASCSKDINSNVVTOGPSOLI FGSKYLHSVQETLSHF  
RAYSLDYSRSTESGAASAFSTSPENITFELDGDSSNSEAGFSGSTFORALEAKYTH  
LDDLQVDDIRSHCVBIHTVIAFAHAETDPQLHTRFALQIVSLYKLNLERICK  
KIISGVSLEKQKDTQSTPHQHLQLQKLRKHQIWRPQGLPKSVSLRNWPF  
QNFLPYPKDSEKHLAIRSGITRSQVSNWPFINARVLKPMIEEMVAEMKRLNNS  
HIQNGPTRLRPKSVWMSQAMHK"  
1423..1453  
/gene="At4g32980"

# gene CDS

304 ATGGACAACAAACAAACAAACAAACAAACAACTTTTAGTTCTCTGGATAATGTCTGACTAACCAA 363  
1 ATGGACAACAAACAAACAAACAAACAAACAACTTTTAGTTCTCTGGATAATGTCTGACTAACCAA 60  
364 AATCCTCTTCTCATGGATTTTATACCTTCAAGAGAAGATTCACTTCATCTTCAACAATG 423  
61 AATCCTCTTCTCATGGATTTTATACCTTCAAGAGAAGATTCACTTCATCTTCAACAATG 120  
424 CTTCCATGGAATACCATCAGATCAGATCAGATCCTCTACAAATGGTGGCTTTTGATATTTTCAAT 483  
121 CTTCCATGGAATACCATCAGATCAGATCCTCTACAAATGGTGGCTTTTGATATTTTCAAT 180  
484 TCTATGCTGACTAACAAATACCTTATCATCTTCTCAAGGTCTATCGATGTTCAAGATAAC 543  
181 TCTATGCTGACTAACAAATACCTTATCATCTTCTCAAGGTCTATCGATGTTCAAGATAAC 240  
544 CGCAATGTTGAGTTCAATGCTCAAT 603  
241 CGCAATGTTGAGTTCAATGCTCAAT 300  
604 TTAAGACATATGATGATTCCTCAAAACAACATGTGGGGTTTTGAAGCAAAATAGTGAGTTT 663  
301 TTAAGACATATGATGATTCCTCAAAACAACATGTGGGGTTTTGAAGCAAAATAGTGAGTTT 360  
664 CAGGCATTTTCAGGTGCTAGTTCGTCCAGTGAACCAATGATGCTTACATTCGGTGAAGAA 723  
361 CAGGCATTTTCAGGTGCTAGTTCGTCCAGTGAACCAATGATGCTTACATTCGGTGAAGAA 420  
724 GATTTCCCGTTTCTAAATTCGAATTAAGAAACAATAGAGCTTTTCATGAGTCTTCGATCA 783  
421 GATTTCCCGTTTCTAAATTCGAATTAAGAAACAATAGAGCTTTTCATGAGTCTTCGATCA 480  
784 GATTTTCTGATGATGCTCGGAGATAAGTCTTTGTGAGCTTACAGATTAGCTCTCAGAG 843  
481 GATTTTCTGATGATGCTCGGAGATAAGTCTTTGTGAGCTTACAGATTAGCTCTCAGAG 540  
844 CAAGCTTCTTGACGACGACAAAGACATTTCTTAACGTTGTTTACTCAAGGTTTCTCTCAA 903  
541 CAAGCTTCTTGACGACGACAAAGACATTTCTTAACGTTGTTTACTCAAGGTTTCTCTCAA 600  
904 CTTATATTTGGCTCAAAATACCTTCACTCTGTTCGAAGAAATACTATCTCATTTTCGCGCA 963  
601 CTTATATTTGGCTCAAAATACCTTCACTCTGTTCGAAGAAATACTATCTCATTTTCGCGCA 660  
964 TACTCGCTCGATTTATTCATCTCGAGAACCGAGTCAAGAGCTGTAGTTACAGCCTTTACT 1023  
661 TACTCGCTCGATTTATTCATCTCGAGAACCGAGTCAAGAGCTGTAGTTACAGCCTTTACT 720  
1024 TCAGTCTTTGAGAATAATACTGAGTTTCTTGATGTTGATTTAACTTCGAGGCGGGT 1083  
721 TCAGTCTTTGAGAATAATACTGAGTTTCTTGATGTTGATTTAACTTCGAGGCGGGT 780  
1084 TTCGGATCTACATTTTCAGAGGAGGAGGATTAGAGCAAGAAACCACTCTCTTTGATCTT 1143  
781 TTCGGATCTACATTTTCAGAGGAGGAGGATTAGAGCAAGAAACCACTCTCTTTGATCTT 840  
1144 CTTCAAAATGGTGGATCGATCGATATAGTCAATTCGCTAGATGAGATTCAACGGTTATATCA 1203  
841 CTTCAAAATGGTGGATCGATCGATATAGTCAATTCGCTAGATGAGATTCAACGGTTATATCA 900  
1204 GCGTTCCATGCTGCACCCAGTTAGTCCAGAGTTACACACCGGTTTGCCTCCCAACC 1263  
901 GCGTTCCATGCTGCACCCAGTTAGTCCAGAGTTACACACCGGTTTGCCTCCCAACC 960  
1264 GTTTCCTCTTATACAAAGAACCTTGAGAGAGAGAACTCTGCAAGAAAGATATCTTCTATGGA 1323  
961 GTTTCCTCTTATACAAAGAACCTTGAGAGAGAGAACTCTGCAAGAAAGATATCTTCTATGGA 1020  
1324 TCTGTTATTCGAGAGAGGCAAGAAAGAACTCTGATCTTATGTTCCAGCAGATTGC 1383  
1021 TCTGTTATTCGAGAGAGGCAAGAAAGAACTCTGATCTTATGTTCCAGCAGATTGC 1080  
1384 CTTCTTTACGAGCTGAAACGAAAGAAACCATCAGATTGAGAGACCTCAACGAGTTTGCT 1443

# 3'UTR

# ORIGIN

Query Match 77.0%; Score 1453; DB 8; Length 1453;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1453; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



763 CTTTCATTGAGTCTTCATCAGATGTTTCTGATGAATGCTCGAGATAAGTCTTTTGCA 822  
 Db 2940 CTTTCATTGAGTCTTCATCAGATGTTTCTGATGAATGCTCGAGATAAGTCTTTTGCA 2999  
 Qy 823 GCTACAAGATTAGCTCAGAGAGCTTCCTGAGCAGCAAGACATTTCTAATAAGTT 882  
 Db 3000 GCTACAAGATTAGCTCAGAGAGCTTCCTGAGCAGCAAGACATTTCTAATAAGTT 3059  
 Qy 883 GTTACTCAAGTTTCTCTCAACTTATTTGGCTCAAAATACCTTCACTCTGTTCAAGAA 942  
 Db 3060 GTTACTCAAGTTTCTCTCAACTTATTTGGCTCAAAATACCTTCACTCTGTTCAAGAA 3119  
 Qy 943 ATACTATCTCAATTCGCGGCATCTGCTCGATTAATTCATCTCAGAGAACCGAGTCAAGAA 1002  
 Db 3120 ATACTATCTCAATTCGCGGCATCTGCTCGATTAATTCATCTCAGAGAACCGAGTCAAGAA 3179  
 Qy 1003 GCTGCTAGTTCAGCTTTTACTTTCAGTTTTCAGATTAATTCATCTCAGAGAACCGAGTCAAGAA 1062  
 Db 3180 GCTGCTAGTTCAGCTTTTACTTTCAGTTTTCAGATTAATTCATCTCAGAGAACCGAGTCAAGAA 3239  
 Qy 1063 TCTAATACTCGAGCGGGTTCGATCTACATTTCAAGAGAGAGCATTAAGAGCAAG 1122  
 Db 3240 TCTAATACTCGAGCGGGTTCGATCTACATTTCAAGAGAGAGCATTAAGAGCAAG 3299  
 Qy 1123 AAAACCCATCTCTCGATCTCTTCAAT----- 1151  
 Db 3300 AAAACCCATCTCTCGATCTCTTCAATGGTATGTAATATCATTCACATTTTTCG 3359  
 Qy 1152 ----- 1151  
 Db 3360 ATAGCTTAAAAATGTTTGATATATATAGTAAGTTTATGATGTTGATATTTCTCTTC 3419  
 Qy 1152 -GGTGGATGATCGATATAGTCTTCGGTAGATGAGATTCATACGGTATATACAGGTTCC 1210  
 Db 3420 AGGTGGATGATCGATATAGTCTTCGGTAGATGAGATTCATACGGTATATACAGGTTCC 3479  
 Qy 1211 ATGCTGCAACCGAGTTAGATTCACAGCTTACACACCGGTTTGCCTCCAAACCGTTTCC 1270  
 Db 3480 ATGCTGCAACCGAGTTAGATTCACAGCTTACACACCGGTTTGCCTCCAAACCGTTTCC 3539  
 Qy 1271 TCTTATACAGAACTGAGAGAGAAATCTGCAAGAGATATCTCTATGGGATCTGTAT 1330  
 Db 3540 TCTTATACAGAACTGAGAGAGAAATCTGCAAGAGATATCTCTATGGGATCTGTAT 3599  
 Qy 1331 TCGAGAGAGCAAGAGAGAGCTCAAGAACTCTATGTTCCACAGAGATTCCTTCTTC 1390  
 Db 3600 TCGAGAGAGCAAGAGAGAGCTCAAGAACTCTATGTTCCACAGAGATTCCTTCTTC 3659  
 Qy 1391 AGCAGCTGAACGAAAGAACCATCAGATTTGAGACCTCAACGAGTTTGCCTGAGAAAT 1450  
 Db 3660 AGCAGCTGAACGAAAGAACCATCAGATTTGAGACCTCAACGAGTTTGCCTGAGAAAT 3719  
 Qy 1451 CTTGTTTCGTTCTACGGAATGATGTTCCAAAGCTTCTTCAACCTTACC 1501  
 Db 3720 CTTGTTTCGTTCTACGGAATGATGTTCCAAAGCTTCTTCAACCTTACC 3770

RESULT 8  
 LOCUS AT26P21/c  
 DEFINITION Arabidopsis thaliana DNA linear pln 16-MAR-2000 project.  
 ACCESSION AL031804  
 VERSION AL031804.1 GI:3688169  
 KEYWORDS Arabidopsis thaliana (thale cress)  
 SOURCE Arabidopsis thaliana  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE 1 (bases 1 to 94482)  
 AUTHORS Bevan, M., Terry, N., Ardiles, W., Buyssehaert, C., Dasseville, R., De

Clerck, R., De Keyser, A., Neyt, P., Rouze, P., Van Den Daele, H., Villard, R., Gielens, J., Van Montagu, M., Hohn, J., Mewes, H. W., Lemcke, K. and Mayer, K. F. X.  
 Unpublished  
 2 (bases 66824 to 67623; 72086 to 72704; 87789 to 88522; 91317 to 92029)  
 Bevan, M., Volckaert, G., Grymonprez, B., Voet, M., Robben, J., Hohn, J., Mewes, H. W., Lemcke, K. and Mayer, K. F. X.  
 Unpublished  
 3 (bases 1 to 94482)  
 EU Arabidopsis sequencing project.  
 Direct Submission  
 Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mpi.biochem.mpg.de, mayer@mpi.biochem.mpg.de  
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk  
 Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.  
 Location/Qualifiers  
 1. 94482  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /variety="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="4"  
 6659..7676  
 /gene="AT4G32890"  
 /join(6659..6877,6969..7676)  
 /gene="AT4G32890"  
 /note="similarity to GATA transcription factor 3, Arabidopsis thaliana, gb:Y13650  
 contains EST gb:AA041075, AI994034.1, AI994545.1, F14085"  
 /codon\_start=1  
 /product="putative protein"  
 /protein\_id="CAA21198.1"  
 /db\_xref="GI:3688170"  
 /db\_xref="GOA:O82632"  
 /db\_xref="SPTREMBL:O82632"  
 /translation="MEKLAPELFLVAGNPDSPVDDLLDFNSDGDGVDGLNTLPDSS TUSTGLTSSNSSLFTDGTGFSGLYFNDDIAEFLWLNKPVESFAGEQDQKLHLF SLGNPQTITGSLTHLIRPELQDFIDDSNVAVPAKARSKRSRAASTWASRL SLADENITPKKQRRVKEQDFAGDMVDVCGSGGRCLCHCATKTKTQWRTPGPGPK TLNACGVRYKSGRLVPEYRSPASPTFVYARHSNHRKVMELRRQKEMRDEHLUSQLR CENLMDIRNSGEDFLMHNNTNHNVPDFRHLI"  
 6659..6877  
 /gene="AT4G32890"  
 /number=1  
 6878..6968  
 /gene="AT4G32890"  
 /number=1  
 6969..7676  
 /gene="AT4G32890"  
 /number=2  
 10392..12039  
 /gene="AT4G32900"  
 /complement(join(10392..10683,10773..10809,10882..10942, 11225..11425,11998..12039))  
 /gene="AT4G32900"  
 /complement(join(10392..10683,10773..10809,10882..10942, 11225..11425,11998..12039))  
 /gene="AT4G32900"  
 /note="Contains Sugar transport proteins signatures AA31-56  
 contains EST gb:AI995200.1, N37475, AA712862"  
 /codon\_start=1  
 /product="putative protein"  
 /protein\_id="CAA21199.1"  
 /db\_xref="GI:3688171"  
 /db\_xref="GOA:O82633"  
 /db\_xref="SPTREMBL:O82633"

JOURNAL  
 REFERENCE  
 AUTHORS  
 JOURNAL  
 REFERENCE  
 TITLE  
 JOURNAL  
 COMMENT  
 FEATURES  
 source  
 gene  
 CDS  
 exon  
 intron  
 exon  
 gene  
 gene  
 CDS

/translation="WPPSSRNSSQPAMRQEKKSISVSFRAENLIPGVVIGFIIGMIL  
DLSSQVTFRRSLSSKVKQSSVPGNKDQELKMWLVVRQLKMRGTGIAQCAH  
AATGYAELMSDRYLLRWEENGQKIWTCXSKSKCKQVKILGDTINKTKHTFYTF  
TTNVKLSLNLKATFISAHENKITEAESVGLPTFFVADAGRTEVCRA"  
exon complement (10392..10683)  
/gene="AT4932900"  
/number=1  
intron complement (10684..10772)  
/gene="AT4932900"  
/number=1  
exon complement (10773..10809)  
/gene="AT4932900"  
/number=2  
intron complement (10810..10881)  
/gene="AT4932900"  
/number=2  
exon complement (10882..10942)  
/gene="AT4932900"  
/number=3  
intron complement (10943..11224)  
/gene="AT4932900"  
/number=3  
exon complement (11225..11425)  
/gene="AT4932900"  
/number=4  
intron complement (11426..11997)  
/gene="AT4932900"  
/number=4  
exon complement (11998..12039)  
/gene="AT4932900"  
/number=5  
gene 12433..17137  
/gene="AT4932910"  
CDS join(12433..12800,12927..13058,13104..13266,13525..13647,  
13691..13747,14081..14281,14381,14491,14596,14633,  
14718..14874,14986..15087,15152..15175,15357..15503,  
15609..15679,15774..15868,16073..16164,16588..16890,  
16934..16974,17077..17137)  
/gene="AT4932910"  
/note="contains EST gb:R89997, AI999140.1"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAA21200.1"  
/db\_xref="GI:3688172"  
/db\_xref="GSPREMBL:O82634"  
/translation="MSSSSGGGLVLFSTKEKTPVLYPLSYGLKSPVHRLSISWGCGN  
NLRTVLNPNELRDDDDGEVGKVVNRLSDEGEISDPQWRAYGVSVPFALLQR  
RNSISLSKMDSSSLQIATWYVYWEYSRDIKSLSNITSLPAPLIEDPRSVIKVRW  
SFWFNWFECCALLMDVPVNIHMQNAEPTSLKAELMELFADTKCLSLPERLVD  
WLSAIEDDPYRNEVMASLSYGMLEI VVSTFFLGLVLCNLILNVKLLHLHGSYLDQ  
LCHRETRRWSQITKLSCSAFWQCAHQTRGLRNMKIMINADCLRAATCNWME  
LFVSHLLYLRPFTKGLDQMSLAQCVOSKPVNTSHKLLRLILGIGENTESVPKNLA  
PGRQAMVAHAMELLTAGSEGEVLVHEQKLGGINMEELHRLVYAQVLSHALTWQ  
TAPIVLACEQKGLLLELLPYRQPVQENQMLIKDLVRSHLISGVHHKHKRGSGI  
FWLQQRDENCLSVIAQQQLFDSVGSLSDESLEKQWGLVELLGSQSGLDLFLHKY  
RDKFSLKVHDGKTIDAAHFAVELRVSCIEAEVELARTFAVCNTQPDVQETTRAI  
NSKREKLAMATALLAVIYASPNRCVISPRIKQISLTGSSSHYQQRGRNHR  
ARYSYSDSDSVLPQDVARLAQTARISLTPAEVLAFSLLSNLLCPTECECTKIRRY  
DWFGLOQVDVNSVEPAIRAG"  
exon 12433..12800  
/gene="AT4932910"  
/number=1  
intron 12801..12926  
/gene="AT4932910"  
/number=1  
exon 12927..13058  
/gene="AT4932910"  
/number=2  
intron 13059..13103  
/gene="AT4932910"  
/number=2  
exon 13104..13266  
/gene="AT4932910"

/number=3  
13267..13524  
/gene="AT4932910"  
/number=3  
13525..13647  
/gene="AT4932910"  
/number=4  
13648..13690  
/gene="AT4932910"  
/number=4  
13691..13747  
/gene="AT4932910"  
/number=5  
13748..14080  
/gene="AT4932910"  
/number=5  
14081..14281  
/gene="AT4932910"  
/number=6  
14282..14380  
/gene="AT4932910"  
/number=6  
14381..14491  
/gene="AT4932910"  
/number=7  
14492..14595  
/gene="AT4932910"  
/number=7  
14596..14633  
/gene="AT4932910"  
/number=8  
14634..14717  
/gene="AT4932910"  
/number=8  
14718..14874  
/gene="AT4932910"  
/number=9  
14875..14985  
/gene="AT4932910"  
/number=9  
14986..15087  
Query Match 58.9%; Score 1110.6; DB 8; Length 94482;  
Best Local Similarity 92.7%; Pred. No. 1e-241;  
Matches 1215; Conservative 0; Mismatches 4; Indels 92; Gaps 1;  
Qy 283 ATAAAGTTTCATGAACCCCAATGGACAAACAACAACAACAACACTTTTACTTCTG 342  
Db 47954 ATGAGTTTCATGAACCCCAATGGACAAACAACAACAACAACACTTTTACTTCTG 47895  
Qy 343 GATAATGTCATGACTAAACCAAAATCCTTCTCATGGATTTTATACCTTCAAGAGAAGAT 402  
Db 47894 GATAATGTCATGACTAAACCAAAATCCTTCTCATGGATTTTATACCTTCAAGAGAAGAT 47835  
Qy 403 TCAACTTCATTCACAATGCTTCCATGGATPACCATCAGATCAGATCCTCTTACAAATG 462  
Db 47834 TCAACTTCATTCACAATGCTTCCATGGATPACCATCAGATCAGATCCTCTTACAAATG 47775  
Qy 463 GGTGGCTTCATATTTTCAATCTTATGCTGACTAACAATACTTATCATCTTCTCCACGG 522  
Db 47774 GGTGGCTTCATATTTTCAATCTTATGCTGACTAACAATACTTATCATCTTCTCCACGG 47715  
Qy 523 TCTATCGATGTTCAAGATAACCGCAATGTTAGTTCATGGCTCCTCCTCTCATCTCTCT 582  
Db 47714 TCTATCGATGTTCAAGATAACCGCAATGTTAGTTCATGGCTCCTCCTCTCATCTCTCT 47655  
Qy 583 CCACTTCATCTTTGGATCATTTTAAGACACTATGATGATTCCTCAACAACATGTGGGT 642  
Db 47654 CCACTTCATCTTTGGATCATTTTAAGACACTATGATGATTCCTCAACAACATGTGGGT 47595  
Qy 643 TTTCGAAGCAATAGTGTAGTTTCAGGCATTTTTCAGGTGTAGTTTGGTCCAAAGTGAACCAATG 702  
Db 47594 TTTCGAAGCAATAGTGTAGTTTTCAGGCATTTTTCAGGTGTAGTTTGGTCCAAAGTGAACCAATG 47535



```
QY 703 ATGCTACATTCGGTGAAGAGATTTCGCCGTTCTTAATTCGATATAAAGAAACAATGAG 762
DB |||||
DB 47534 ATGCTACATTCGGTGAAGAGATTTCGCCGTTCTTAATTCGATATAAAGAAACAATGAG 47475
QY 763 CTTTCATTGAGTCTTCGATCAGATGTTTCTGATGAATGCTCGGAGATAAGTCTTTGTGCA 822
DB |||||
DB 47474 CTTTCATTGAGTCTTCGATCAGATGTTTCTGATGAATGCTCGGAGATAAGTCTTTGTGCA 47415
QY 823 GCTACAAGATTAGCTCAGACAGCTTCCTGACAGCAGCAAGACATTTCTAATAAGGTT 882
DB |||||
DB 47414 GCTACAAGATTAGCTCAGACAGCTTCCTGACAGCAGCAAGACATTTCTAATAAGGTT 47355
QY 883 GTTACTCAAGGTTTCTCTCAACTATATTGCGCTCAAAATACCTTCACCTCTGTTCAAGAA 942
DB |||||
DB 47354 GTTACTCAAGGTTTCTCTCAACTATATTGCGCTCAAAATACCTTCACCTCTGTTCAAGAA 47295
QY 943 ATACTATCTCATTTCCGCGCATCTCGCTCGATTAATCATCTCGAGGAACCGAGTCAAGG 1002
DB |||||
DB 47294 ATACTATCTCATTTCCGCGCATCTCGCTCGATTAATCATCTCGAGGAACCGAGTCAAGG 47235
QY 1003 GCTGCTAGTTCAGCCTTTTACTTCAGGTTTTCAGATATAAAGTCTTTGATGGTGAT 1062
DB |||||
DB 47234 GCTGCTAGTTCAGCCTTTTACTTCAGGTTTTCAGATATAAAGTCTTTGATGGTGAT 47175
QY 1063 TCTAATAACTCGGAGCGGGTTTCGGATCTACATTTCAAAGGAGAGCATTTAGAGCAAG 1122
DB |||||
DB 47174 TCTAATAACTCGGAGCGGGTTTCGGATCTACATTTCAAAGGAGAGCATTTAGAGCAAG 47115
QY 1123 AAAACCCATCTCTGGATCTCTTCAAT----- 1151
DB |||||
DB 47114 AAAACCCATCTCTGGATCTCTTCAATGGTATGTAATATATTCATTTCACTTTTTCG 47055
QY 1152 ----- 1151
DB 47054 ATAAGTTAAAAAATGGTTTGTATATATATGATGAAGTTTATGATGATATTTCTCTTC 46995
QY 1152 -GGTGGATGATGATATAGTCAATTCGGTAGATGAGATTCATACGGTTATATACCGTTC 1210
DB |||||
DB 46994 AGGTGGATGATGATATAGTCAATTCGGTAGATGAGATTCATACGGTTATATACCGTTC 46935
QY 1211 ATGCTCAACCGAGTTAGATCCACAGTTACACACCGGTTTGCCTCCAAACCGTTTCC 1270
DB |||||
DB 46934 ATGCTCAACCGAGTTAGATCCACAGTTACACACCGGTTTGCCTCCAAACCGTTTCC 46875
QY 1271 TCTTATACAAGAACTCTGAGAGAGAGAAATCTCGAAGAGATTAATCTCTATGGGATCTGTAT 1330
DB |||||
DB 46874 TCTTATACAAGAACTCTGAGAGAGAGAAATCTCGAAGAGATTAATCTCTATGGGATCTGTAT 46815
QY 1331 TGGAGAGAGGCAACACAGACTCAAGAACTCTATGTTCCACAGACTTGCCTTCTTC 1390
DB |||||
DB 46814 TGGAGAGAGGCAACACAGACTCAAGAACTCTATGTTCCACAGACTTGCCTTCTTC 46755
QY 1391 AGCAGCTGAACGAAAGAACCATCATGATTTGGAGACCTCAACGAGGTTTGCCTGAGAAAT 1450
DB |||||
DB 46754 AGCAGCTGAACGAAAGAACCATCATGATTTGGAGACCTCAACGAGGTTTGCCTGAGAAAT 46695
QY 1451 CTGTTTCGGTTCTACGGAATGGATGTTCCAAACTTCCTTCACCCCTTACC 1501
DB |||||
DB 46694 CTGTTTCGGTTCTACGGAATGGATGTTCCAAACTTCCTTCACCCCGTAAC 46644
```

```
RESULT 9
ATCHRIV78/c
LOCUS ATCHRIV78 198669 bp DNA linear PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 78.
ACCESSION AL161582
VERSION AL161582.2 GI:7270218
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
```

```
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 38818)
Benes, V., Rechmann, S., Borkova, D., Ansoorge, W., Mewes, H.W.,
Lemcke, K. and Mayer, K.F.X.
Unpublished
JOURNAL
REFERENCE
AUTHORS
2 (bases 38710 to 90803)
Lennard, N., Quail, M., Harris, B., Rajandream, M.A., Barrell, B.G.,
Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
Unpublished
JOURNAL
REFERENCE
AUTHORS
3 (bases 90697 to 195178)
Terry, N., Ardiles, W., Buyssehaert, C., Dasseville, R., De Clerck, R.,
De Keyser, A., Neyt, P., Rouze, P., Van Den Daele, H., Villarroel, R.,
Gielens, J., Van Montagu, M., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
Unpublished
JOURNAL
REFERENCE
AUTHORS
4 (bases 147462 to 198669)
Peters, S.A., van Staveren, M., Dirkse, W., Stiekema, W., Mewes, H.W.,
Lemcke, K. and Mayer, K.F.X.
Unpublished
JOURNAL
REFERENCE
AUTHORS
5 (bases 157520 to 158319; 162782 to 163400; 178485 to 179218;
182013 to 182725)
Volckaert, G., Grymonprez, B., Voet, M., Robben, J., Mewes, H.W.,
Lemcke, K. and Mayer, K.F.X.
Unpublished
JOURNAL
REFERENCE
AUTHORS
6 (bases 1 to 198669)
EU Arabidopsis sequencing, project.
Direct Submission
Biochemie, Am Klopferstr. 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mpibiochem.mpg.de, mayer@mpibiochem.mpg.de
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7JU Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRIV77 at the 5' end and an
overlap with ATCHRIV79 at the 3' end.
FEATURES
Location/Qualifiers
1. 198669
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
9071..11457
/gene="AT4G32720"
/join(9071..9121,9225..9303,9403..9552,9830..9972,
10018..10211,10325..10447,10537..10831,10937..11084,
11145..11262,11307..11457)
/gene="AT4G32720"
/feature="similarity to RNA-binding protein LAH1,
Saccharomyces cerevisiae, PIR2:548600
contains EST gb:F15295"
/codon_start=1
/product="putative protease"
/protein_id="CAB79989.1"
/db_xref="GI:7270219"
/db_xref="GOA:O65529"
/db_xref="SPTREXBL:O65529"
/translation="MSIFCLTEETAKTVLRQVEFFSDSNLPIDDFLAKTTESEDL
VSLALICFSKMRGKYLKSGDDIPEDTAKADVTLSALKIDSGKVKGRSTE
LLKLEDLIELNARTVAASPFSDYDKREDVDFSSQYKIGYTSFWMYSSLOVNSV
RMRHVAESRIFSGVALVEFTEADQNMKNLQVAGLELFPKKEFDNREKDEY
KANTOPQSGSNKQNSDHNKNSAYEPDYPKGLISFTLKRSAEGTEQKSEPT
DNTMESKTPADTADKENTGVQASGAEDEDEKEKGALATHKDKNVDVREDL
KAVFKGQDVKFDKMGSTCYLRFDEPEASQARAALANEGGLAVKRFVLEP
VIVKSLVIALCOAKLRSLGSLSEAKIGLTRVAVEGEFFCLVLQIILVLYAGE
EEEEGGSGERGQIHVVAGTNLRWKKHFRFS"
9071..9121
/gene="AT4G32720"
/number=1
9122..9224
/gene="AT4G32720"
exon
intron
```

exon /number=1  
9225..9303  
/gene="AT4g32720"  
/number=2  
intron 9304..9402  
/gene="AT4g32720"  
/number=2  
exon 9403..9552  
/gene="AT4g32720"  
/number=3  
intron 9553..9829  
/gene="AT4g32720"  
/number=3  
exon 9830..9972  
/gene="AT4g32720"  
/number=4  
intron 9973..10017  
/gene="AT4g32720"  
/number=4  
exon 10018..10211  
/gene="AT4g32720"  
/number=5  
intron 10212..10324  
/gene="AT4g32720"  
/number=5  
exon 10325..10447  
/gene="AT4g32720"  
/number=6  
intron 10448..10536  
/gene="AT4g32720"  
/number=6  
exon 10537..10831  
/gene="AT4g32720"  
/number=7  
intron 10832..10936  
/gene="AT4g32720"  
/number=7  
exon 10937..11084  
/gene="AT4g32720"  
/number=8  
intron 11085..11144  
/gene="AT4g32720"  
/number=8  
exon 11145..11262  
/gene="AT4g32720"  
/number=9  
intron 11263..11306  
/gene="AT4g32720"  
/number=9  
exon 11307..11457  
/gene="AT4g32720"  
/number=10  
gene 12797..16010  
/gene="AT4g32720"  
join(12797..12867,13004..13058,13283..13340,13507..13601,  
13702..13805,13902..14030,14192..16010)  
/gene="AT4g32730"  
/notes="similarity to transforming protein A-myb, Xenopus  
laevis, PIR1:S36095  
Contains Myb DNA-binding domain repeat signatures  
A29-37;Myb DNA-binding domain repeat signatures  
A481-89;Myb DNA-binding domain repeat signatures A449-72  
contains EST gb:AW004163.1"  
/codon\_start=1  
/product="putative myb-protein"  
/protein\_id="CA879990.1"  
/db\_xref="GI:7270220"  
/db\_xref="GGA:Q9S7G7"  
/db\_xref="SPTREMBL:Q9S7G7"  
/translation="MKREMKAPTTPLSLQGLKKGQRTSGPARSTKGQWTPPEDE  
VLCKAVERFQGNKKIAECFDRIDVQCLHRWQVLNPELVKGPWSKEEDNTIIDLV  
EKYGPKKWTISGHLPGRIKOCRRWHLNPKNAWTOEBELTILRAHQIYGNK  
WAEIMKFLGRSDNSIKHNWSSVKKLDSYVASGLLDQCQSSPLIALQNKSIASSSS

WMISNGDESSRFGVDAERSECSQASTVFSQSTNDLQDEVQORNEEYMPFHSCTEQ  
QISNAASHAPYYPFKDKIVVPEISCEETCSKFKQNLNCSHELRTTATDLPFGV  
SNDKQDRGLELLTHNMDGKNQALQQDFQSSVRLSDQFLNSDTPPEAQTITDE  
ECCRVLFPDNNKDSSTSSGEQGRNVDPNQKSLCSQAAETHAHTGKVPALPWHPS  
SSEGLAGNCVPLLLDSDLPRNDSNAPIQGRLFGATELECKTDTNDGFDITYG  
HVTSHGNDNGGPFQOGLSYIPKDSLKLVPLNFSFSPSRVNIYFIDDKPAEKDKG  
ALCYEPRFSAADIPFSCDLVFSNDDLQEVSPFGIRQLMISSMNCVPLRLWDSPC  
HDRSPVNLMDTAKSFGAPSIILKKRRHLLSPVLDKRDKKLKRATATSLANDFRRL  
DVMLDEGDDCMTRSPESPEDKNICASPSIARDNRNCASARLYQEWIPIDEEPKEITLE  
SGGVTSQMNQNGNDGASAKNVSPSLHLIIWYQL"  
exon 12797..12867  
/gene="AT4g32730"  
/number=1  
intron 12868..13003  
/gene="AT4g32730"  
/number=1  
exon 13004..13058  
/gene="AT4g32730"  
/number=2  
intron 13059..13282  
/gene="AT4g32730"  
/number=2  
exon 13283..13340  
/gene="AT4g32730"  
/number=3  
intron 13341..13506  
/gene="AT4g32730"  
/number=3  
exon 13507..13601  
/gene="AT4g32730"  
/number=4  
intron 13602..13701  
/gene="AT4g32730"  
/number=4  
exon 13702..13805  
/gene="AT4g32730"  
/number=5  
intron 13806..13901  
/gene="AT4g32730"  
/number=5  
exon 13902..14030  
/gene="AT4g32730"  
/number=6  
intron 14031..14191  
/gene="AT4g32730"  
Query Match 58.9%; Score 1110.6; DB 8; Length 198669;  
Best Local Similarity 92.7%; Pred. No. 9.9e-242; Gaps 1;  
Matches 1215; Conservative 0; Mismatches 4; Indels 92; Gaps 1;  
Qy 283 ATAAAGTTTCATAGAAACCCCAATGGACACAAACAAACAAACACTTTTAGTTCTGT 342  
Db 138650 ATGCAGTTTCATAGAAACCCCAATGGACACAAACAAACAAACACTTTTAGTTCTGT 138591  
Qy 343 GATAATGTCATGACTACCAAAATCCTTCTCTCATGGATTTTATACCTTCAAGGAGAGAT 402  
Db 138590 GATAATGTCATGACTACCAAAATCCTTCTCTCATGGATTTTATACCTTCAAGGAGAGAT 138531  
Qy 403 TCAACTTCATCTCAACAATGCTTCCATGGAATACCATCAGATCAGATCCTCTTACAAATG 462  
Db 138530 TCAACTTCATCTCAACAATGCTTCCATGGAATACCATCAGATCAGATCCTCTTACAAATG 138471  
Qy 463 GGTGGCTTGTATATTTTCAATTTCTATGCTGACTGAACAAATACATTATCATCTTCTCCACGG 522  
Db 138470 GGTGGCTTGTATATTTTCAATTTCTATGCTGACTGAACAAATACATTATCATCTTCTCCACGG 138411  
Qy 523 TCTATCGATGTTCAAGATAACCGCAATGTTGAGTTCTATGGCTCTCTCTCTCTCATCTCTCCT 582  
Db 138410 TCTATCGATGTTCAAGATAACCGCAATGTTGAGTTCTATGGCTCTCTCTCTCTCATCTCTCCT 138351  
Qy 583 CCACCTCATCTCTTGGATCAATTTTAAGACATPATCATGATCTCTCTCTCTCTCTCTCTCTCTCTCT 642  
Db 138350 CCACCTCATCTCTTGGATCAATTTTAAGACATPATCATGATCTCTCTCTCTCTCTCTCTCTCTCTCT 138291

QY	643	TTTGAAGCAAAATAGTGGATTTTCCAGGATTTTCCAGGTGATGTTGGTCCAAGTGAACCAATG	702
Db	138290	TTTGAAGCAAAATAGTGGATTTTCCAGGATTTTCCAGGTGATGTTGGTCCAAGTGAACCAATG	138231
QY	703	ATGCTACATTCGGTGAAGAAATTTCCCGTTTCTAATTTTGAATTAAGAAACCAATGAG	762
Db	138230	ATGCTACATTCGGTGAAGAAATTTCCCGTTTCTAATTTTGAATTAAGAAACCAATGAG	138171
QY	763	CTTTCAATTCAGTCTTCCATCAGATGTTTCTGATGAATGCTCCGAGATAAGTCTTTGTGCA	822
Db	138170	CTTTCAATTCAGTCTTCCATCAGATGTTTCTGATGAATGCTCCGAGATAAGTCTTTGTGCA	138111
QY	823	GCTACAGATTAAGCTTCAGAGCAAGCTTCTTCAGAGCAAGAAAGATTTCTAATAACGTT	882
Db	138110	GCTACAGATTAAGCTTCAGAGCAAGCTTCTTCAGAGCAAGAAAGATTTCTAATAACGTT	138051
QY	883	GTCTACTCAAGGTTTCTCTCAACTTATTTGCTCAAAATACCTTCACTCTGTTCAAGAA	942
Db	138050	GTCTACTCAAGGTTTCTCTCAACTTATTTGCTCAAAATACCTTCACTCTGTTCAAGAA	137991
QY	943	ATACTATCTCAATTCGCCCATACTCGCTCGATTATTCATCTCGAGGAACCGAGTCAGGA	1002
Db	137990	ATACTATCTCAATTCGCCCATACTCGCTCGATTATTCATCTCGAGGAACCGAGTCAGGA	137931
QY	1003	GGTGTAGTTTACGCTTACCTTACGTTTGTGATATTAACCTGATTTCTTGTGATGAT	1062
Db	137930	GGTGTAGTTTACGCTTACCTTACGTTTGTGATATTAACCTGATTTCTTGTGATGAT	137871
QY	1063	TCTAATACTCGAGCGGGTTTCGGATCTCATTTCAAAGGAGAGCATTAAGAACCAAG	1122
Db	137870	TCTAATACTCGAGCGGGTTTCGGATCTCATTTCAAAGGAGAGCATTAAGAACCAAG	137811
QY	1123	AAACCCATCTCTTGATCTTCTTCAAT	1151
Db	137810	AAACCCATCTCTTGATCTTCTTCAAT	137751
QY	1152	-----	1151
Db	137750	ATAAGTTAAAAAATGGTTTGATATATATGATGAAGTTTATGAGTTCATATTTCTCTTC	137691
QY	1152	-GGTGAATGATCGATATAGTCAATTCGGTAGATGAGATTCATACGGTTATATCAGCGTTCC	1210
Db	137690	AGGTGAATGATCGATATAGTCAATTCGGTAGATGAGATTCATACGGTTATATCAGCGTTCC	137631
QY	1211	ATGCTGCACCCAGTTAGATCCACAGTTACACACCGGTTTGCCTCCAAACCGTTTCCCT	1270
Db	137630	ATGCTGCACCCAGTTAGATCCACAGTTACACACCGGTTTGCCTCCAAACCGTTTCCCT	137571
QY	1271	TCTTATACAAAGAACCTTGAGAGAGAGATCTGCAAGAGATAATCTCTATGGGATCTGTAT	1330
Db	137570	TCTTATACAAAGAACCTTGAGAGAGAGATCTGCAAGAGATAATCTCTATGGGATCTGTAT	137511
QY	1331	TGGAGAGGCAAGACAGACAGTCAAGAACTCTCTATGTTCCACAGCATTCGCTTCTTC	1390
Db	137510	TGGAGAGGCAAGACAGACAGTCAAGAACTCTCTATGTTCCACAGCATTCGCTTCTTC	137451
QY	1391	AGCAGTCAAAACGAAAGAACCAATCATGATTTGGAGACCTCAACGAGGTTTGGCTGAGAAAT	1450
Db	137450	AGCAGTCAAAACGAAAGAACCAATCATGATTTGGAGACCTCAACGAGGTTTGGCTGAGAAAT	137391
QY	1451	CTGTTTCGGTTCTACGGAATGCGATGTTCCAAACCTTCCTTCAACCTTACC	1501
Db	137390	CTGTTTCGGTTCTACGGAATGCGATGTTCCAAACCTTCCTTCAACCTTACC	137340

RESULT 10  
 AT283320  
 LOCUS  
 DEFINITION A.thaliana partial ATH1 gene.  
 ACCESSION 283320  
 VERSION 283320.1 GI:1865680  
 KEYWORDS ATH1 gene; MAR.

SOURCE	Arabidopsis thaliana (thale cress)
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE	1 (bases 1 to 10072) van Drunen,C.M., Oosterling,R.W., Keultjes,G.M., Weisbeek,P.J., van Driel,R. and Smeekens,S.C.
TITLE	Analysis of the chromatin domain organisation around the plastocyanin gene reveals an MAR-specific sequence element in Arabidopsis thaliana
JOURNAL	Nucleic Acids Res. 25 (19), 3904-3911 (1997)
MEDLINE	98026209
PUBMED	9380515
REFERENCE	2 (bases 1 to 10072) Oosterling,R.W.
AUTHORS	Direct Submission
TITLE	Submitted (11-DEC-1996) Oosterling R.W., Utrecht University, Molecular cell Biology, Padualaan 8, Utrecht, The Netherlands, 3584 CH
FEATURES	Location/Qualifiers
source	1..10072 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /variety="Columbia-0" /db_xref="taxon:3702" /chromosome="4" /clone="PH1pr" /clone_lib="Lambda ZAP genomic library"
misc_feature	1..1431 /note="ATH1 locus MAR-1" /evidence=experimental
misc_signal	471..489 /note="ATH1 MAR-1 MAR-specific sequence signature" /evidence=experimental
misc_feature	2046..3762 /note="ATH1 locus MAR-2" /evidence=experimental /evidence=experimental
misc_signal	7905..9103 /note="ATH1 locus MAR-3" /evidence=experimental
gene	9388..10072 /gene="ATH1"
CDS	9388..10072 /gene="ATH1" /note="Related sequences: ATH1 genomic clone, acc. no. X80127; cDNA ATH1 clone, acc. no. X80126" /codon_start=1 /protein_id="CAB05910.1" /db_xref="GI:1865681" /db_xref="GOA:P48731" /db_xref="SWISS-PROT:P48731" /translation="MDNNNNNTFSSLDNVTNQNPPLMDTFPSREDSTSFSTMLPWN TTRSPLOWGGDIENSMLTKYLSRPSIDVDNVRVEMAPPPHPPHPLDHLR HYDSSNNWGFANSEFQAGVGVGPSPMSTFGEEDFFPLISKNKNNELSLAS DVSDECSISLCLATRLASEQASCSKDISNNVVTQGFSLIFGSKYLHSVQELIUSHP AAYSIDYSSR"
ORIGIN	Query Match 37.3%; Score 702.8; DB 8; Length 10072; Best Local Similarity 99.7%; Pred. No. 4.8e-149; Matches 704; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	283 ATAAAGTTTCATAGAAACCAATGAGCAACAAACAAACAAACACTTTAGTTCTCTG 342
Db	9367 ATGCAGTTTCATAGAAACCAATGAGCAACAAACAAACAAACACTTTAGTTCTCTG 9426

QY	343	GATAATGTCATGACGTAACCAAAATCTCTCTCTCATGGATTTTATACCTTCAAGAGAAGAT	402
Db	9427	GATAATGTCATGACGTAACCAAAATCTCTCTCTCATGGATTTTATACCTTCAAGAGAAGAT	9486
QY	403	TCAACTTTCATCTCAACAATGCTTCCATGGAATACCATCAGATCAGATCCTCTCAAAATG	462
Db	9487	TCAACTTTCATCTCAACAATGCTTCCATGGAATACCATCAGATCAGATCCTCTCAAAATG	9546
QY	463	GGTGGCTTTGATATTTTCAATTTCTATGCTGATCAAAATATCTTATCATCTTCTCCACGG	522
Db	9547	GGTGGCTTTGATATTTTCAATTTCTATGCTGATCAAAATATCTTATCATCTTCTCCACGG	9606
QY	523	TCTATCGATGTTCAAGATAACCGCAATGTTAGTTTCATGGCTCCTCCCTCCATCCTCT	582
Db	9607	TCTATCGATGTTCAAGATAACCGCAATGTTAGTTTCATGGCTCCTCCCTCCATCCTCT	9666
QY	583	CCACTTTCATCTTTGGATCATTTAAGACACTATGATGATTCCTCAAAACAATGCGGT	642
Db	9667	CCACTTTCATCTTTGGATCATTTAAGACACTATGATGATTCCTCAAAACAATGCGGT	9726
QY	643	TTTGAAGCAAAATGAGCTTTAGGCAATTTTCAAGTGATGTTGGTCCCAAGTCAACCAATG	702
Db	9727	TTTGAAGCAAAATGAGCTTTAGGCAATTTTCAAGTGATGTTGGTCCCAAGTCAACCAATG	9786
QY	703	ATGCTTACATTCGGTGAAGAAGATTTCCCGTTTCTAAATTTCAAAATAAAAGAAACAATGAG	762
Db	9787	ATGCTTACATTCGGTGAAGAAGATTTCCCGTTTCTAAATTTCAAAATAAAAGAAACAATGAG	9846
QY	763	CTTTCAATGAGCTTGCATCAGATGTTTCTGATGATGCTCGGAGATGATGCTTTGTGCA	822
Db	9847	CTTTCAATGAGCTTGCATCAGATGTTTCTGATGATGCTCGGAGATGATGCTTTGTGCA	9906
QY	823	GCTACAAGATTAGCTTCAGACGAAGCTTCTTCAGCAGCAAAAGACATTTCTAATAACGTT	882
Db	9907	GCTACAAGATTAGCTTCAGACGAAGCTTCTTCAGCAGCAAAAGACATTTCTAATAACGTT	9966
QY	883	GTTACTCAAGGTTTCTCTCAACTTATTTGGCTCAAAATACCTTCACTCTGTTCAAGAA	942
Db	9967	GTTACTCAAGGTTTCTCTCAACTTATTTGGCTCAAAATACCTTCACTCTGTTCAAGAA	10026
QY	943	ATACATCTCATTTGCCGCATACCTCGCTCGATTATTCATCTCGAG	988
Db	10027	ATACATCTCATTTGCCGCATACCTCGCTCGATTATTCATCTCGAG	10072
RESULT 11			
LOCUS	AF375967 1182 bp mRNA linear PLN 02-JUN-2003		
DEFINITION	Lycopersicon esculentum bell-like homeodomain protein 4 (BL4) mRNA, partial cds.		
ACCESSION	AF375967		
VERSION	AF375967.1 GI:31323448		
KEYWORDS	Lycopersicon esculentum (tomato)		
SOURCE	Lycopersicon esculentum		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon. 1 (bases 1 to 1182)		
REFERENCE	Ron N., Parnis A. and Lifschitz, E.		
AUTHORS	Direct Submission		
JOURNAL	Submitted (03-MAY-2001) Biology, Technion, Haifa, Israel		
FEATURES	Location/Qualifiers		
source	1..1182		
	/organism="Lycopersicon esculentum"		
	/mol_type="mRNA"		
	/db_xref="taxon:4081"		
gene	<1..1182		
	/genes="BL4"		
CDS	<1..973		
	/genes="BL4"		
	/note="TBL4"		
	/codon_start=2		

		/product="bell-like homeodomain protein 4"	
		/protein_id="AAP47026.1"	
		/db_xref="GI:31323448"	
		/translation="RDQCSDIISGSGVTHAFQRRFDSELTSCNSRNLSLSPGSKYKPV YLSOFLTGSRYLWMOEILSEIAQLSLQNHNLVGYRGNGTENGANT'SFALNSDAGRY AAMSDDSPGLMGCBAKKNLVALLQVDDQVNOCLDETHWISAFHAYTELDPSIH ARPALOTISLSYKNLGENKQSHSRNGRTPQORMRKSGSEVILKHSFKSSGSHFOOL KRKHQJWRFPQGLPERSVSLRAWMFQNFLEHYPKDAEKQLLAVSKGLTRQSUNWF INARVLWKPMI BEMVAEMNRKIRAGNEEDHRNHKILIESHLFTMK"	
Query Match		13.5%; Score 255; DB 8; Length 1182;	
Best Local Similarity		61.8%; Pred. No. 2.8e-47;	
Matches 521; Conservative 0; Mismatches 230; Indels 32; Gaps 5;			
QY	826	ACAAGATTAGCTCAGAGCAAGCTTTTGCAGCAGCAAAAGACATTTCTAATAACGTTGTT	885
Db	59	AGACGCTTTGATTTCAGAGTTAACTTCTTGCACAGCAGAAATCTGCTTTAAGCTTTGTT	118
QY	886	-----ACTCAAGGTTTCTCTCAACTTATATTTCGCTCAAAATACCTTCACTCTGTT	936
Db	119	TCATACAAACCTGTCTACCTTTTCAAAATTTCTTACTGGATCAAGATATCTCCGTGTAATG	178
QY	937	CAAGAAATACATCTCAATTTTGGCGCATATCGCTCGATTTATTCATCTCAGAGCAACGAG	996
Db	179	CAAGAAATACATCTCTGAAATTCGTCAGTTATCACTTCAAAATCATAACTTGTAGGCTAT	238
QY	997	TCAGAGCTGCTA-GTTCAGCCTTACTTTCAGCTTTTGAGATATATACGATGTTCTTGA	1055
Db	239	CGAGAAATGGAGCGGAGAAATGGTGCAACACTTTCATTTGCTTTGAAATTCAGATCTGGA	298
QY	1056	TGGTGATTCTTAATAACTTCGGAGCGGGTTTCGGATCTACATTTCAAAGGAGCAATTAGA	1115
Db	299	AGGGCT-----ATGCCACCATGAGTTCTGATGATTCACCTGATGTTCTGATGGTTGTA	354
QY	1116	AGCAAGAAACCACTCTCTTGGATCTCTTCAAAATGGTGGATGATCGATATAGCTATTG	1175
Db	355	AGCTAAGAAAGAAATCTGTAGCTTTGTCGAAGTGGTTGACGATCAATACAAACCAATG	414
QY	1176	CGTAGATGAGATTCATACGGTTTATCAGCGTTTCCATGCTGCAACCGAGTTAGATCCACA	1235
Db	415	CTTAGATGAGATTCATATGATGTCATCTCTGATTTTCATCTGCTGACCGAGTTGGATCTAG	474
QY	1236	GTTACACACCGGTTTCCCTCCAAACCGTTTCTTCTTATACAAAGAACCTTGA-GAGAGA	1294
Db	475	TATACATGCTCGTTTCGCCCTTCAACAATCTCTCTCTGTACAAAAACCTTGAGGGGAGA	534
QY	1295	GAATCTCAGAGAGATTAATCTCTATGGGATCTGTATTGGAGAGAGCAAGACAAAGACTC	1354
Db	535	GAATAAGCAATCATCTCTCGAAATGGAGAAACATTTCAACAAGAGATGAGAGAAAGGG	594
QY	1355	AA-----GAAACCTCTATGTTTCCACCAGCA-TTGCCTTCTTTCAGCAGCT	1397
Db	595	GAGTAGAGAGTCAATTTTGAACATCATTTCAAAAGCAGTGGCACTTTCAGCAGCT	654
QY	1398	GAAACGAAAGAACCATCAGATTTGGAGACCTCAACGAGGTTTGCTCAGAAATCTGTTTC	1457
Db	655	AAAAACGAAAGATCATCAGTTATGGAGGCCACAAAGAGGTTTGCAGAAAGATCTGTATC	714
QY	1458	GTTTCTACGGAATTTGGATGTTCCAAAACTTCTTCACTTACCTTACCCGAAAGATTCGGAGAA	1517
Db	715	GTTTTGAGCATGGATGTTTCAAACTTCTTCACCCATATCCAAAGATGAGAGAA	774
QY	1518	ACATCTCTAGCTATACAGAGTGGCTTGAACAGAGTCAGGTATCAACTGGTTTATAAA	1577
Db	775	GCAATTGCTAGCAGTAAAAAGTGGATTGACAAAGAGTCAGGTATCGAATTTGTTTCAAAA	834
QY	1578	TGCGCGGTTTAGGCTATGGGAAGCCGATGATAGAGAGATGTATCGGAAATGAACAAGAG	1637
Db	835	TGCTCGATTCCTCTATGGAACCAATGATAGAGAAATGTATGCTGAGATGAACAGAG	894
QY	1638	GAA	1640

```

Db      895 AAA 897

RESULT 12
AF406703      2065 bp mRNA linear PLN 14-JUL-2003
LOCUS Solanum tuberosum BEL1-related homeotic protein 30 (Bel30) mRNA,
DEFINITION complete cds.
ACCESSION AF406703
VERSION AF406703.1 GI:22652126
KEYWORDS Solanum tuberosum (potato)
SOURCE Solanum tuberosum
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 2065)
AUTHORS Chen,H., Rosin,F.M., Prat,S. and Hannapel,D.J.
TITLE Interacting transcription factors from the three-amino acid loop
extension superclass regulate tuber formation
JOURNAL Plant Physiol. 132 (3), 1391-1404 (2003)
PUBMED 12857821
REFERENCE 2 (bases 1 to 2065)
AUTHORS Chen,H., Rosin,F.M. and Hannapel,D.J.
TITLE A KNOX protein of potato interacts with several members of the TALE
family of transcription factors
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 2065)
AUTHORS Chen,H., Rosin,F.M. and Hannapel,D.J.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-2001) Department of Horticulture, Iowa State
University, 257 Horticulture Hall, Ames, IA 50011-1100, USA
FEATURES
Source
1. .2065
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Desiree"
/db_xref="taxon:4113"
1. .2065
/gene="Bel30"
39. .1976
/gene="Bel30"
/Note="StBel30"
/codon_start=1
/product="BEL1-related homeotic protein 30"
/protein_id="AA03627.1"
/db_xref="GI:22652127"
/translation="MATYFPSPNNORDADQTFQYFROSLPSYSEASNAPENMMVFN
YSSGAYSDMLTGTSSQOHCIDIPSIGATPNTSQOELISNLGSGOMGIDFSSWD
SENEMLADNVQAVQVQGLSLGSLNSIPSGTGISHVQSNPQGGFNFPGDD
NSQPEQKNAYFPFDPNGRDLDMKGNYSPTGSSARTIPSSKYLKAAQYLLEDEV
VONKLAUSLMLDEIRYQYHQMOIVVSSFDVAVAGEAKPYTALALQTSRHPF
CLRDAICQIRASRSLGEQASNSKAIGISRLRFVDDHRIQORALQQLGMQOHAW
RPQRLPSSVSLRAWLFEHLHPYKDSKIMLARQTLGTRQVSNWFINARVLW
KPMVEMKKEBAGDAKIDNSSSDVAPRLAKDSKVERGELHONAAEFEOYNSGOI
LESKNHEADVMEGASNAETQSGKVENOTGEPLPAMDNCTLFDAFVQSGNPFSEF
SGFSGNVLPNGVSLTGLQOGBSGLNPLMSIETHSVVPLRADDMYSTATTNVPETA
EFNCLDSGNRQOPFWLLPSAT"

ORIGIN
Query Match 7.0%; Score 132; DB 8; Length 2065;
Best Local Similarity 70.2%; Pred. No. 2,4e-19;
Matches 177; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1375 CAGCATTCCTTTCAGCAGCTGAAACGAAGAACCATCAGATTGGAGACCTCAGGA 1434
|||||
DB 1167 CAGCAGAGAGCCCTCAGCAGCTTGGTATGATGCAACAACTGCGAGCCCTCAGAGG 1226
|||||

QY 1435 GGTTCGCTGAGAAATCTGTTTCGGTTCTACGGAATGCGATGTTCCAAAACCTCTTCAC 1494
|||||
DB 1227 CGATTGCTGAGAGCTCTGTTTCAGTTTGGTGTCTTGGCTCTTGGACACTTCTTCAT 1286
|||||

```

```

QY 1495 CCTTACCCGAAAGATTCCGGAGAACATCTTCTAGCTATACGAGTGGCTTGACAGAACT 1554
|||||
DB 1287 CCTTACCCGAAAGATTCTTACAAAATATGCTAGCAAGCAAACTGGCTTTACGAGAGT 1346
|||||
QY 1555 CAGGTATCAAACTGGTTTATAAATGCGCGGTGTAGCTATGGAAGCGCATGATAGAGAG 1614
|||||
DB 1347 CAGGTATCAAACTGGTTTATAAATGCAACGGTGTGCTTTTGGAAACCACTGTTAGGAA 1406
|||||
QY 1615 ATGTATCGCGAA 1626
|||||
DB 1407 ATGTACAAAGAA 1418
|||||

RESULT 13
AF406697      2735 bp mRNA linear PLN 14-JUL-2003
LOCUS Solanum tuberosum BEL1-related homeotic protein 5 (Bel5) mRNA,
DEFINITION complete cds.
ACCESSION AF406697
VERSION AF406697.1 GI:22652114
KEYWORDS Solanum tuberosum (potato)
SOURCE Solanum tuberosum
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 2735)
AUTHORS Chen,H., Rosin,F.M., Prat,S. and Hannapel,D.J.
TITLE Interacting transcription factors from the three-amino acid loop
extension superclass regulate tuber formation
JOURNAL Plant Physiol. 132 (3), 1391-1404 (2003)
PUBMED 12857821
REFERENCE 2 (bases 1 to 2735)
AUTHORS Chen,H., Rosin,F.M. and Hannapel,D.J.
TITLE A KNOX protein of potato interacts with several members of the TALE
family of transcription factors
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 2735)
AUTHORS Chen,H., Rosin,F.M. and Hannapel,D.J.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-2001) Department of Horticulture, Iowa State
University, 257 Horticulture Hall, Ames, IA 50011-1100, USA
FEATURES
Source
1. .2735
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Desiree"
/db_xref="taxon:4113"
1. .2735
/gene="Bel5"
148. .2214
/gene="Bel5"
/Note="StBel5"
/codon_start=1
/product="BEL1-related homeotic protein 5"
/protein_id="AA03621.1"
/db_xref="GI:22652115"
/translation="WYQGTSDNTNIQADHQORHNHNSNNNIOTLYLMPNNYMQG
YTTSDTQQOQLFLNSSFPAASNALCHANIQAPLQOQHFVGVPLPFAVSLDQINHHG
LIQRMNNQDQQQVIVPSSVTSVATCGGTTDLASQAFORPITPCHRAQQQQQ
GLSLSLPQLOQIISFNINISSSPRTNITRTGLDSSNMVLGSKYLKAAQLLD
EVNIVGSKIQGDQKDNMKNESPLASDVNTNNSGGSSSSKQKNEAVELTTAQ
ROELCKDAKILAMLEEVQRYOYHOMIIVLSEFQVAGISAKSYQALHAISK
QFQGLPERAVSLRAHLFEHLHPYKDSKIMLARQTLGTRQVSNWFINARVL
WIPWEMLEEVKQENSTNGNKNKNTISAPNEEKHPITITSLQQGITTQ
ABISTSTISPTGASLHAHNFSLGSFNMNDNTTVDHIENNAKKORMKHFSF
SSILSSVMEAKARESSNKGFTNPLMAAYAMGDFQFDPHQQMTANFHNGVSLT
GLPPSENLAAMPVSQQNYLNDLSRSEMGSHYMRMYENIDFQSGNKEFFPTQLLPDFV
TGNLGT"

ORIGIN
Query Match 6.8%; Score 128.6; DB 8; Length 2735;

```



Sakazume, N., Sano, H., Sasaki, D., Sato, K., Sato, K., Satoh, K., Shibata, K.,  
Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,  
Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami, T., Takeda, Y.,  
Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,  
Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,  
Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and  
Yoshimura, A.  
Direct Submission  
Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of  
Agrobiological Sciences, Department of Molecular Genetics, Head of  
Laboratory of Gene Expression; 2-1-2 Kanondai, Tsukuba, Ibaraki  
305-8602, Japan E-mail: skikuchi@nias.affrc.go.jp,  
Tel: 81-29-838-7007, Fax: 81-29-838-7007  
This clone is one of the 28K full-length cDNA clones from japonica  
rice.

URL : http://cdna01.dna.affrc.go.jp/cDNA/  
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,  
Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,  
Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,  
Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and  
Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y.,  
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,  
Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,  
Mizuno, K., Nakikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S.,  
Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,  
Yoshimura, A., Matsubara, K. and Murakami, K.  
Genome Exploration Research Group in Riken Genomic Sciences Center  
and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,  
Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanacaki, T.,  
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,  
Hirooka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,  
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,  
Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,  
Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,  
Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,  
Yasunishi, A. and Hayashizaki, Y.

#### FEATURES

Location/Qualifiers  
1..2570  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="J013124105"

#### ORIGIN

Query Match	6.7%	Score 126.4;	DB 8;	Length 2570;
Best Local Similarity	68.4%	Pred. No. 4.5e-18;		
Matches 175;	Conservative 0;	Mismatches 81;	Indels 0;	Gaps 0;

```

QY 1375 CAGCATTCCTTCTTCAGACGTGAACACGAAGAACCATCAGATTGGAGACCTCAACGA 1434
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1482 CAACAGCGCGCTTCCAGCAGTATGGTTTGTACAGCAAAATGCTGGAGGCCACAGAGG 1541

QY 1435 GGTTCGCTGAGAAATCTGTTTCGTTCTACGGAATGGATGTTCCAAACTTCCTTCAC 1494
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1542 GGACTGCCCGAAATCACTGATTCATCTTCGTCTGGCTTTTGAACACTTCCTTCAC 1601

QY 1495 CTTTACCCGAAGATTCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAGAAGT 1554
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1602 CGGTATCCAAAGATTCAGAAAGCTGATGCTAGCGACAAACTGGCTTAACAAGAAGT 1661

QY 1555 CAGGTATCAACTGTTTATTAATCGCGGCTTAGGCTATGGAGCGCATGATAGAGAG 1614
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1662 CAGATTTCAATGTTTTCATAAATGCCCGTGTCCGCTGTGGAAACCGATGATCGAAGAC 1721

QY 1615 ATGTATCGGAAATGA 1630
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1722 ATGTATAAAGAAGACA 1737
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 2, 2004, 08:25:17 ; Search time 113.5 Seconds  
(without alignments)  
9390.037 Million cell updates/sec

Title: US-09-423-575-1  
Perfect score: 3284  
Sequence: 1 attagttataaaagtgtgc.....ttaaaaaaaaaaaaaaaaaa 1886

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delopt 6.0 , Delext 7.0

Searched: 1566107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlh  
-O=/cpn2.1/USPTO.spool/US09423575/runat.01092004.161105.18548/app.query.fasta\_1.2055  
-DB=A\_Geneseq.25Jan04 -QFMT=fasten -SUFFIX=rag -MINMATCH=0.1 -LOOPTCL=0  
-LCOEPT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09423575@CGN.1.1.145@runat.01092004.161105.18548 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq.29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2484	75.6	473	5 AAU76514	Arabidops
2	2484	75.6	473	5 AAU76885	Arabidops
3	472	14.4	680	6 ADB23150	Enviro
4	472	14.4	680	7 ADD30304	Plant yle
5	444	13.5	514	3 AAG14898	Arabidops
6	444	13.5	524	3 AAG14897	Arabidops
7	444	13.5	524	4 AA502524	Arabidops
8	444	13.5	533	3 AAG14896	Arabidops
9	424	12.9	739	3 AA502510	Arabidops
10	424	12.9	739	7 ADE37297	Plant yle

11	419	12.8	459	3 AAG14140	Arabidops
12	419	12.8	459	3 AAG32446	Arabidops
13	419	12.8	482	3 AAG32445	Arabidops
14	419	12.8	482	3 AAG14139	Arabidops
15	419	12.8	491	3 AAG21086	Arabidops
16	419	12.8	539	3 AAG21085	Arabidops
17	415	12.6	418	3 AAG21087	Arabidops
18	412.5	12.6	610	4 AAE02469	Arabidops
19	412.5	12.6	610	7 ADE37117	Plant yle
20	411.5	12.5	433	3 AAG32447	Arabidops
21	411.5	12.5	433	3 AAG14141	Arabidops
22	408	12.4	627	4 AAE02525	Arabidops
23	408	12.4	627	7 ADC46619	Thalecres
24	408	12.4	627	7 ADD30787	Plant yle
25	403	12.3	471	3 AAG50499	Arabidops
26	403	12.3	494	3 AAG50498	Arabidops
27	395.5	12.0	435	3 AAG50500	Arabidops
28	379	11.5	598	3 AAG21938	Arabidops
29	379	11.5	598	3 AAG21937	Arabidops
30	377	11.5	608	5 ABB79411	Hordeum v
31	376.5	11.5	611	5 ABB79412	Hordeum v
32	348.5	10.6	500	3 AAG21939	Arabidops
33	348.5	10.6	612	6 AAO27114	QSH-1 gen
34	348.5	10.6	612	6 AAO27115	QSH-1 gen
35	338	10.3	153	3 AAB32901	Pinus rad
36	332	10.1	176	3 AAB32619	Eucalyptu
37	327.5	10.0	107	3 AAB32960	Pinus rad
38	326	9.9	126	3 AAB32644	Eucalyptu
39	301	9.2	233	5 ABB60937	Novel flo
40	299	9.1	202	3 AAG08089	Arabidops
41	243	7.4	293	3 AAG50529	Arabidops
42	243	7.4	316	3 AAG50528	Arabidops
43	235.5	7.2	257	3 AAG50530	Arabidops
44	232	7.1	126	3 AAB32653	Eucalyptu
45	222	6.8	86	3 AAB32893	Pinus rad

ALIGNMENTS

RESULT 1  
AAU76514  
ID AAU76514 standard; protein; 473 AA.  
XX  
AC AAU76514;  
AC  
DT 05-JUN-2002 (first entry)  
DE  
DE Arabidopsais Athl protein.  
XX  
KW Grass; plant; herbicide resistance; grass flowering inhibitor; baseball;  
KW cricket; football; golf; rugby; soccer; tennis; lawn; park;  
KW athletic field; animal feedstuff; Athl protein.  
XX  
OS Arabidopsais thaliana.  
XX  
PN WO200214524-A2.  
XX  
PD 21-FEB-2002.  
XX  
PF 16-AUG-2001; 2001WO-EP009572.  
XX  
PR 18-AUG-2000; 2000US-0226422P.  
PR 27-NOV-2000; 2000US-0253274P.  
PR 27-NOV-2000; 2000US-0253327P.  
PR 22-JUN-2001; 2001US-0300220P.  
XX  
(ADVA-) ADVANTA SEEDS BV.  
XX  
PI Van Der Valk P, Van Dun CMP, Smeekens SCM, Proveniers MCG;  
XX  
DR WPI; 2002-257611/30.  
DR N-FSDS; ABK11039.  
XX

PT New genetically modified grasses that express inhibited generative  
PT propagation, or herbicide resistance, useful for forage (e.g. cattle  
PT feedstuff) or amenity purposes (e.g. for use in an athletic field, lawn  
PT or park).

XX Example 1; Fig 6; 56pp; English.

XX The invention relates to a grass plant, which has been genetically  
CC modified to substantially inhibit generative propagation and carry  
CC herbicide resistance. The grass is useful for growth and/or propagation  
CC of grasses. The grass is used to play at least one sport (e.g. baseball,  
CC cricket, football, golf, rugby, soccer or tennis), or used at least in a  
CC portion of an athletic field, lawn or park. The grass is fed to animal  
CC (e.g. cattle, goat, horse or sheep) or used as an animal feedstuff. The  
CC present sequence represents the amino acid sequence of Arabidopsis Ath1  
CC protein, which was used to inhibit flowering in grasses

XX Sequence 473 AA;

Alignment Scores:  
Pred. No.: 5,45e-250 Length: 473  
Score: 2484.00 Matches: 473  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 75.64% Indels: 0  
DB: 5 Gaps: 0

US-09-423-575-1 (1-1886) x AAU76514 (1-473)

QY 304 ATGGACAAACAAACAAACAACTTTTCTCGATAATGTCAGCTAACCAA 363  
DB 1 MetAspAsnAsnAsnAsnAsnAsnThrPheSerSerLeuAspAsnValMetThrAsnGln 20  
QY 364 AATCTCTCTCATGGATTTTATACCTTCAAGAGAAGATTCACTTCACTTCTCAACATG 423  
DB 21 AsnProLeuLeuMetAspPheLeuProSerArgGluAspSerThrSerPheSerThrMet 40  
QY 424 CTTCCATGATPACCATCAGATCCTCTCAAAATGGGTGGCTTTGATATTTTCAAT 483  
DB 41 LeuProTrpAsnThrIleArgSerAspProLeuGlnMetGlyGlyPheAspIlePheAsn 60  
QY 484 TCTATGCTGACTAAACAATACTATCATCTCTCCACGGTCTATCGATGTTCAAGATAAC 543  
DB 61 SerMetLeuThrAsnLysTyLeuSerSerSerProArgSerIleAspValGlnAspAsn 80  
QY 544 CCAATGTTGAGTTTCATGGCTCCTCTCTCATCTCTCCATCTCATCTTTCGGATCAT 603  
DB 81 ArgAsnValGluPheMetAlaProProProHisProProProLeuHisProLeuAspHis 100  
QY 604 TTAAGACACTATGATGATTCCTCAACAACTATGGGGGTTTGAAGCAATAGTGGATTT 663  
DB 101 LeuArgHisTyAspAspSerSerAsnAsnMetTrpGlyPheGluAlaAsnSerGluPhe 120  
QY 664 CAGGCATTTTCAGGTGATGTTGGTCCAAAGTGAACCAATGATGTTTACATTCGGTGAAGAA 723  
DB 121 GlnAlaPheSerGlyValValGlyProSerGluProMetMetSerThrPheGlyGluGlu 140  
QY 724 GATTTCCCGTTTCTTAATTCGAATAAAGAAACATAGCTTTTCAATGAGCTTTGATCA 783  
DB 141 AspPheProPheLeuIleSerAsnLysArgAsnAsnGluLeuSerLeuSerLeuAlaSer 160  
QY 784 GATGTTTCTGATGATGTCGGAGATAGTCTTTCTCGAGTCAAGATAGCTTCAGAG 843  
DB 161 AspValSerAspGlySerGluLeuSerLeuCysAlaAlaThrArgLeuAlaSerGlu 180  
QY 844 CAAGCTTCTTCAGCAGCAAGACATTTCTAATAACGTTGTACTCAAGGTTTCTCTCAA 903  
DB 181 GlnAlaSerCysSerSerLysAspIleSerAsnAsnValValThrGlnGlyPheSerGln 200  
QY 904 CTTATATTTGGCTCAAAATACCTTCACTCTTCAAGAAATACTATCTCATTTCGCCGA 963  
DB 201 LeuIlePheGlySerLysTyLeuHisSerValGlnGlnLeuIleLeuSerHisPheAlaAla 220

QY 964 TACTCCTCGATTATTCATCTCTGAGGAACCGAGTCAGGAGCTGCTAGTTCAGCCTTTACT 1023  
DB 221 TyrSerLeuAspTySerSerArgGlyThrGluSerGlyAlaAlaSerSerAlaPheThr 240  
QY 1024 TCACGTTTTGAGAAATAAATACTGAGTTTCTTGATGGTGATTCTTAATAACTCGAGGGGGT 1083  
DB 241 SerArgPheGluAsnIleThrGluPheLeuAspGlyAspSerAsnAsnSerGluAlaGly 260  
QY 1084 TTCGGATCTACATTTTCAAGGAGAGCATTTAGAGCAAGAAACCAACCATCTCTTGATCTT 1143  
DB 261 PheGlySerThrPheGlnArgAlaLeuGluAlaLysLysThrHisLeuLeuAspLeu 280  
QY 1144 CTTCAATGGTGGATGATGATATAGTCTTCCGGTAGATGAGATTCATACAGGTTTATCA 1203  
DB 281 LeuGlnMetValAspAspArgTySerHisCysValAspGluIleHisThrValIleSer 300  
QY 1204 GCCTTCCATGCTCCAAACCGAGTTAGATCCACAGTTTACACACCCGGTTTGCCTCCAAACC 1263  
DB 301 AlaPheHisAlaAlaThrGluLeuAspProGlnLeuHisThrArgPheAlaLeuGlnThr 320  
QY 1264 GTTTCCTTCTTATACAGAACCTGAGAGAGAGAATCTCCAGAGATATCTCTATGGGA 1323  
DB 321 ValSerPheLeuTyThrLysAsnLeuArgGluArgIleCysLysLysIleIleSerMetGly 340  
QY 1324 TCTGTATTGAGAGAGGCAAGACAAAGACTCAAGAAACCTCTATCTTCCACCAGCATTCG 1383  
DB 341 SerValLeuGluArgGlyLysAspLysThrGlnGlnThrSerMetPheHisGlnHisCys 360  
QY 1384 CTTCTTCAGAGCTGAAACGAAAGAACCATCAGATTTGGAGACCTCAACGAGGTTTGCT 1443  
DB 361 LeuLeuGlnGlnLeuLysArgLysAsnHisGlnIleTrpArgProGlnArgGlyLeuPro 380  
QY 1444 GAGAAATCTGTTTCGGTCTCTCGGAATGGATGTTCCAAACCTTCTTCCACCTTACCCTG 1503  
DB 381 GluLysSerValSerValLeuArgAsnTrpMetPheGlnAsnPheLeuHisProTyPro 400  
QY 1504 AAAGATTCGAGAAACATCTTCTAGCTATACGAATGGCTTGACAAAGTCAAGTATCA 1563  
DB 401 LysAspSerGluLysHisLeuLeuAlaIleArgSerGlyLeuThrArgSerGlnValSer 420  
QY 1564 ACTCGTTTATAAATCGCGGCTAGGCTATCGAGCCGATGATAGAGAGATGATGCG 1623  
DB 421 AsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetIleGluGluMetIlyAla 440  
QY 1624 GAAATGAACAGAGGAAGCTCAATACAGTCACTTCAACCCACGACCACTTCTCGA 1683  
DB 441 GluMetAsnLysArgLysLeuAsnAsnSerHisIleGlnProAsnGlyProThrLeuArg 460  
QY 1684 ATGCCAAATCTGTTATGATGAGCCCAAGCAATGCATAA 1722  
DB 461 MetProLysSerValMetMetSerGlnAlaMetHisLys 473

RESULT 2

AAU76885  
ID AAU76885 standard; protein; 473 AA.

AC AAU76885;  
XX

DT 05-JUN-2002 (first entry)

DE Arabidopsis thaliana homeobox gene Ath1 encoded protein.

XX Grass; plant; transgenic; flowering inhibition; inflorescence;  
KW tiller production; delayed heading; gibberellic acid; phytohormone; golf;  
KW genetically modified grass; athletic field; sport; baseball; cricket;  
KW football; rugby; soccer; tennis; lawn; landscaping; cattle; horse; sheep;  
KW goat; animal feedstuff; Ath1; homeobox gene.

OS Arabidopsis thaliana.

XX WO200214486-A2.

XX 21-FEB-2002.



QY 1684 ATGCCAAAATCTGTTATGATGAGCCAAGCAATGCATAAA 1722  
Db 461 MetProLysSerValMetMetSerGlnAlaMetHisLys 473

### RESULT 3

ADB23150 standard; protein: 680 AA.

Qy	514	TCGCCGGTCTATCGATGTTCAAGATAACCGCAATGTTGAGTTCATGGCTCTCTCTCTCT	573
Db	91	GlualaAlaSerIleThrAlaAlaAspAsnIleSerValIeuHisGlyIlePyrProPheArg	110
Qy	574	CATCCTCCTCCACTTCATCCTTTGGATCAATTAAAGACACTATGATTCCTCAAAACAC	633
Db	111	ValGlnTyrSerLeuTyrGly-----SerHisGlnValAspProThrHisGln	126
Qy	634	ATGTGGGGTTTTGAAGCAATAAGTAGTTTCAGGCAATTTTCAGGTGTAGTTCGCAAGT	693
Db	127	GlnAlaAlaCysGluThrProArgAlaGlnGlnGlyLeuSerLeuThrLeuSerSerGln	146
Qy	694	GAACCAATGATGCTACATTCGGTGAAGAAGATTTCCCGTTCTTAATTCGATATAAAGA	753
Db	147	GlnGlnGlnGlnGlnHisHisGlnGlnHisGlnProIleHisValGlyPheGlySer	166
Qy	754	AACAATCAGCCTTCATTGAGTCGTGATCAGATGTTTCTGATCAATGCTCGGAGATAAGT	813
Db	167	GlyHisGlyGluAspIleArgValGlySer-----	176
Qy	814	CTTTGTGCACCTACAAGATTAGCCTCAGACAAGCTCTTCGCAGCAAGACATTTCT	873
Db	177	-----	179
Qy	874	AATAACGTTGTTACTCAAGGTTTCTCTCAACTTATATTTGGCTCAAAATACCTTTCACT	933
Db	180	GlySerGlyValThrSrnglyIleAlaAsnLeuVal--SerSerIleTyrLeuLysAla	198
Qy	934	GTTCAAGAATACTACTCATTTCCGCCGATACCTCGCTCGATTATTCATCTCGAGGAACC	993
Db	199	AlaGlnGluLeuLeuAspGluValValAsnAlaAspSerAsp-----	212
Qy	994	GAGTCAGGAGCTCTAGTTGAGCCTTTACTTTCACGTTTTGAGAAATATA--ACTGAGTTT	1050
Db	213	AspMetAsnAlaLysSerGlnLeuPheSerSerLysGlySerCysGlyAsnAspLys	232
Qy	1051	CTTGATGGTGAATCTTAATACTCGGAGCGGGTTTCGGATCT-----	1092
Db	233	ProValGlyGluSerSerAlaGlyAlaGlyGlyGluGlySerGlyGlyAlaGluAla	252
Qy	1093	-----ACATTCAAGGAGAGCATTAGAACAAGAA	1125
Db	253	AlaGlyLysArgProValGluLeuGlyThrAlaGluArgGlnGluIleGlnMetLysLys	272
Qy	1126	ACCCATCTCTTGATCTCTTCAAATGGTGGATGATCGATATGATTCGCTAGATGAG	1185
Db	273	AlaLysLeuSerAsnMetLeuHisGluValGluGlnArgTyrArgGlnTyrHisGlnGln	292
Qy	1186	ATTCATACGGTTATATCAGCGTTCCATGCTGCACCCGAGTTA--GATCCACAGTTACAC	1242
Db	293	MetGlnMetValIleSerSerPheGluGlnAlaAlaGlyIleGlySerAlaLysSerTyr	312
Qy	1243	ACCCGGTTGGCCTCCAAACCGTTCTCTTATACAGAACCTCGACAGAGAACTCTGC	1302
Db	313	ThrSerLeuAlaLeuLysThrIleSerArgGlnPheArgCysLeuLysGluAlaIleAla	332
Qy	1303	AGAAGATATCTCTATGGATCTGTATTTGGACAGAGGCAACACAGACTCAAGAAACC	1362
Db	333	-----GlyGlnIleLysAlaAlaAsnLysSerLeuGlyGluGluAsp	346
Qy	1363	TCTATG-----TTCCACGAGATTCGCTT	1386
Db	347	SerValSerGlyValGlyArgPheGluGlySerArgLeuLysPheValAspHisHisLeu	366
Qy	1387	CTTCAGCAG-----CTGAACGAAAGAACCATCAGATTTCG	1422
Db	367	ArgGlnGlnArgAlaLeuGlnGlnLeuGlyMetIleGlnHisProSerAsnAsnAlaP	386
Qy	1423	AGACCTCAACGAGGTTGCTCAGAAATCTGTTTCGGTCTCAGGAAATGGATGTTCCAA	1482
Db	387	ArgProGlnArgGlyLeuProGluArgAlaValSerValLeuArgAlaTrpLeuPheGlu	406
Qy	1483	AACTTCCTCTCACCCCTTACCAGAAAGATTCGGAAACATCTCTAGCTATACGAATGCC	1542

```

:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
407 HisPheLeuHisProTyrProLysAspSerAspLysHisMetLeuAlaLysGlnThrGly 426
1543 TTGCACGAAGTCAAGTATCAAACTGGTTTATAAAATCGCGGGTTAGGCTATGGAACCG 1602
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
427 LeuThrArgSerGlnValSerAsnTppHeileAsnAlaArgValArgLeuTyrLysPro 446
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1603 ATCATAGACAGATGATCGGGAATGACACAGACGAAGCTCAATTAACAGTCACATTCAA 1662
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
447 MetValGluGluMetTyrMetGluGluMetLysGluGlnAlaLysAsnMetGlySerMet 466
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1663 CCCAACGGACCA-----ACTCTTCGAATG 1686
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
467 GluLysThrProLeuAspGlnSerAsnGluAspSerAlaSerLysSerThrSerAsnGln 486
1687 CGAAATCTGTTATGATGACCGCAAGCAATGCATAAATATAGACAACAAT 1734
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
487 GluLysSerProMetAlaAspThrAsnTyrHisMetAsnProAsnHis 502
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

RESULT 4
ADD30304
ID ADD30304 standard; protein; 680 AA.
XX ADD30304;
XX
XX 15-JAN-2004 (first entry)
XX
XX Plant yield-related protein from clone G1589.
XX
XX transcription factor; transgenic plant; growth rate; senescence;
XX seed germination rate; plant vigor; seedling vigor.
XX
XX Arabidopsis thaliana.
XX
XX WO2003013227-A2.
XX
XX 20-FEB-2003.
XX
XX 09-AUG-2002; 2002WO-US025805.
XX
XX 09-AUG-2001; 2001US-0310847P.
XX 19-NOV-2001; 2001US-033649P.
XX 11-DEC-2001; 2001US-0338692P.
XX 14-JUN-2002; 2002US-00171468.
XX
XX (MEND-) MENDEL BIOTECHNOLOGY INC.
XX
XX Patcliffe O, Riechmann JL, Adam LJ, Dubell AT, Heard JE;
XX Pilgrim ML, Jiang C, Reuber TL, Creelman RA, Pineda O, Yu G;
XX Brown PE;
XX
XX WPI; 2003-248221/24.
XX
XX N-PSDB; ADD30303.
XX
XX New plant transcription factor polynucleotides and polypeptides, useful
XX in producing transgenic plants with commercially valuable properties,
XX such as an alteration in a plant growth characteristic, e.g. growth rate
XX or apomixis.
XX
XX Disclosure; SEQ ID NO 333; 454pp; English.
XX
XX The invention relates to a number of isolated Arabidopsis thaliana cDNA
XX sequences and their encoded proteins which are especially transcription
XX factor related cDNA's and proteins. The isolated or recombinant plant
XX transcription factor polynucleotides and polypeptides are useful in
XX producing transgenic plants with commercially valuable properties, i.e.
XX modified or altered desirable traits as compared to a reference plant,
XX such as an alteration in a plant growth characteristic, e.g. growth rate,
XX germination rate of seeds, vigor of plants and seedlings, or leaf and
XX flower senescence. Sequence information related to the polynucleotides
XX and polypeptides can also be used in bioinformatic search methods. The
XX transgenic plant is useful for growing a progeny plant from a parent
XX plant. This sequence represents one of the proteins of the invention.

```



```

Length:      514
Matches:     163
Conservative: 74
Mismatch:    170
Indels:      150
Gaps:        21

```



[illegible]

Qy	1723	-----TAAGACAACAATTGTGTTTACCACACTTTGTGATAATTA	176
Db	437	rSerGlnGlnAsnGlnGlyAsnAsnAsnAniileProTyThrSerAsp-AlaC	457
Qy	1761	GGCAAATGCTACTCTATCATGTGCCCAAAACCTAAACCATGTACGACT	1807
Db	457	lucGlnAsnLeuValPheAlaAspProLysProaspArgalaThrThr	472
 RESULT 6 AAG14897 ID AAG14897 standard; protein; 524 AA. XX AC AAG14897; XX XX DT DT XX XX DE 17-OCT-2000 (first entry) XX Arabidopsis thaliana protein fragment SEQ ID NO: 14932. KW Protein identification; signal transduction pathway; metabolic pathway; KW hybridisation assay; Genetic mapping; gene expression control; promoter; KW termination sequence. XX XX OS Arabidopsis thaliana. XX XX PN EP1033405-A2. XX XX PD 06-SEP-2000. XX XX PF 25-FEB-2000; 2000BP-00301439. XX XX PR 25-FEB-1999; 99US-0121825P; PR 05-MAR-1999; 99US-0123180P; PR 03-MAR-1999; 99US-0123548P; PR 23-MAR-1999; 99US-0125788P; PR 25-MAR-1999; 99US-0126264P; PR 29-MAR-1999; 99US-0126785P; PR 01-APR-1999; 99US-0127462P; PR 08-APR-1999; 99US-0128234P; PR 08-APR-1999; 99US-0128714P; PR 16-APR-1999; 99US-0129845P; PR 19-APR-1999; 99US-0130077P; PR 21-APR-1999; 99US-0130449P; PR 23-APR-1999; 99US-0130510P; PR 28-APR-1999; 99US-0130891P; PR 30-APR-1999; 99US-0131449P; PR 30-APR-1999; 99US-0132048P; PR 04-MAY-1999; 99US-0132407P; PR 05-MAY-1999; 99US-0132484P; PR 06-MAY-1999; 99US-0132485P; PR 06-MAY-1999; 99US-0132486P; PR 07-MAY-1999; 99US-0132487P; PR 11-MAY-1999; 99US-0132863P; PR 14-MAY-1999; 99US-0134218P; PR 14-MAY-1999; 99US-0134219P; PR 14-MAY-1999; 99US-0134221P; PR 14-MAY-1999; 99US-0134370P; PR 18-MAY-1999; 99US-0134768P; PR 19-MAY-1999; 99US-0134941P; PR 20-MAY-1999; 99US-0135124P; PR 21-MAY-1999; 99US-0135353P; PR 24-MAY-1999; 99US-0135629P; PR 25-MAY-1999; 99US-0136021P; PR 27-MAY-1999; 99US-0136392P; PR 28-MAY-1999; 99US-0136782P; PR 01-JUN-1999; 99US-0137222P; PR 03-JUN-1999; 99US-0137528P; PR 04-JUN-1999; 99US-0137502P; PR 07-JUN-1999; 99US-0137724P; PR 08-JUN-1999; 99US-0138094P; PR 10-JUN-1999; 99US-0138540P; PR 14-JUN-1999; 99US-0138647P; PR 14-JUN-1999; 99US-0139116P;			

```

PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 20-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147203P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 16-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155859P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 21-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161320P.
PR 28-OCT-1999; 99US-0161322P.
PR 28-OCT-1999; 99US-0161393P.
PR 29-OCT-1999; 99US-0162142P.

```

Alignment Scores:  
 Pred. No.: 8,42e-37  
 Score: 444.00  
 Percent Similarity: 42.55%  
 Best Local Similarity: 29.26%  
 Query Match: 13.52%  
 DB: 3

Length: 524  
 Matches: 163  
 Conservative: 74  
 Mismatches: 170  
 Indels: 150  
 Gaps: 21



PA (RATC/) RATCLIFFE O.  
 PA (CREE/) CREELMAN R.  
 PA (JIANG/) JIANG C.  
 PA (PINE/) PINEDA O.  
 PA (REUB/) REUBER L.  
 PA (ADAM/) ADAM L.  
 XX  
 PI Heard J, Ratcliffe O, Creelman R, Jiang C, Pineda O, Reuber L;  
 PI Adam L;  
 XX  
 DR WP1; 2001-335978/35.  
 DR N-PSDB; AAD06493.  
 XX  
 PT Nucleic acids encoding plant transcription factor polypeptides, useful  
 PT for altering the pathogen resistance characteristics of plants, e.g.  
 PT corn, potato and cotton plants.  
 XX  
 XX  
 XX Claim 4; Page 112-114; 134pp; English.  
 CC The present sequence is Arabidopsis thaliana transcription factor  
 CC homologue. The transcription factors are used to alter the structure and  
 CC developmental characteristics of plants such as soybean, wheat, corn,  
 CC potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf,  
 CC banana, blackberry, blueberry, strawberry, raspberry, carrot, cantaloupe,  
 CC cauliflower, cucumber, coffee, eggplant, grapes, mango, lettuce,  
 CC honeydew, melon, onion, papaya, peppers, pineapple, spinach, squash,  
 CC sweet corn, tobacco, tomato, peas, watermelon, roseaceous fruits and  
 CC vegetable brassicas. The transcription factors are specifically useful  
 CC for modifying traits associated with plant's pathogen tolerance such as  
 CC alterations in cell wall composition, trichome number or structure,  
 CC callose induction, phytoalexin induction, and alterations in the cell  
 CC death response. Transgenic plants expressing these transcription factors  
 CC are more tolerant to biotrophic or necrotrophic pathogens such as fungi,  
 CC bacteria, molluscs, viruses, nematodes and parasitic higher plants. The  
 CC transcription factors are also used in gene therapy  
 XX  
 SQ Sequence 524 AA;

Alignment Scores:  
 Pred. No.: 8 42e-37 Length: 524  
 Score: 444.00 Matches: 163  
 Percent Similarity: 42.53% Conservative: 74  
 Best Local Similarity: 29.26% Mismatches: 170  
 Query Match: 13.52% Indels: 150  
 DB: 4 Gaps: 21

US-09-423-575-1 (1-1886) x AAE02524 (1-524)

QY 303 ATGGACAAACACAAACAAACACATTTAGTTCTCTGGAT----- 345  
 DB 22 AsnGluGlnGlnGlnGlnGln-AlaSerSerSerSerAlaAlaSerPheSerGl 41  
 QY 346 -----AATGTCATGACTAAACCAAAATCCTCTCTCATGGATTTTATACCTTC 392  
 DB 41 uIleValSerGlyAspValArgAsnAsnGlu-----MetValPheIleProPr 57  
 QY 393 AAGAGAAGATCAACTTATCTCAACAATGTTCCATGGAATACCATCAGATCATCC 452  
 DB 57 othrSerAspValAlaValAsnGlyAsnVal-----ThrValSerSerAsnAs 73  
 QY 453 TCTACAAATG-----GGTGGCTTTGATATTTTCAATCTATGCTGACTACAAATACTT 506  
 DB 73 pheuSerPheHisGlyGlyGlyLeuSerLeuSerLeuGlyAsnGlnIleGlnSerAla 93  
 QY 507 ATCATCTTCTCCACGGTCTATCGATGTTTCAAGATAACCGCAATGTTGAGTTTCATGGCTCC 566  
 DB 93 lserValSerProPheGlnTyr----- 100  
 QY 567 TCCTCTCATCTCTCTCCACTTCATCTTGGATCATTTAAGACACTATGATGATCTCTC 626  
 DB 101 -----HisTyrGlnAsnLeuSe 106  
 QY 627 AAACACATGTGGGTTTTTGAAGCAAATAGTAGTGTTCAGGCATTTTCAGGTTAGTTGG 686

116 rAsnGlnLeuSerTyrAsnAsnLeuAsnPro----- 116  
 QY 687 TCCAAGTGAACCAATGATGCTACTTCGGTGAAGAAGATTTCCCGTTTCTAATTTTCGAA 746  
 DB 117 -----SerThrMetSerAspGlu----- 122  
 QY 747 TAAAGAAACAANTGAGCTTTCATGAGTCTTGATCAGATGTTTCTGATGAATGCTCGGA 806  
 DB 123 -----AsnGlyLysSerLeuSerValHisGlnHisSerAspGlnIleLeuPr 139  
 QY 807 GATAAGTCTTTGTGCAGCTACAAGATTAGCCTCAGACGAAGCTTCTGCGACGACCAAGA 866  
 DB 139 oSerSerVal-----TyrAsnAs 145  
 QY 867 CATTTCTAATAACGTTGTT-----ACTCAAGGTTTCTC 899  
 DB 145 nAsnGlyAsnAsnGlyValGlyPheTyrAsnAsnTyrArgTyrGluThrSerGlyPheVal 165  
 QY 900 TCAACTATATATTTGGTCAAAATACCTCTACTCTGTTCAAGTAATACTATCTCATTTCCG 959  
 DB 165 lSerSerValLeuArgSerArgTyrLeuLysProThrGlnGlnLeuLeuAspGluValva 185  
 QY 960 CGCATACTCGCTCGATTAT-----TCATCTCGAGAACCGA 995  
 DB 185 lSerValArgLysAspLeuLysLysMetLysAsnAspLysGlyGlnAs 205  
 QY 996 GTCAGGAGCTGCTAGTTTACGCCTTTACTTACCGTTTGTGAGATATTAATCAGTTTCTGA 1055  
 DB 205 pPheHisAsnGlySerSer-----AspAsnIleThrGlu-----As 217  
 QY 1056 TGGTGATTCTTAATTAATTCGAGGGGGGTTTCGGATCTACATTTCAAAGAGAGACATTAGA 1115  
 DB 217 pAspLysSerGlnSerGlnGlnLeu-----SerProSerGluArgGlnGlnLeuGl 234  
 QY 1116 AGCAAGAAACCCATCTCTTGGATCTTCTTCAAAATGGTGGATGATGATATAGTCTATTG 1175  
 DB 234 nSerLysLysSerLysLeuLeuThrMetValAspGluValAspLysArgTyrAsnGlnTy 254  
 QY 1176 CGTAGATGAGATTCATACGGTTTATATCAGGTTCCATGCTGCAACCGAGTTAGAT---CC 1232  
 DB 254 rHisHisGlnMetGluAlaLeuAlaSerSerPheGluMetValThrGlyLeuGlyAlaAl 274  
 QY 1233 ACAGTTACACACCCGGTTTCCCTCCAAACCGTTTCTCTTTATATACAAGAACCTTGAGAGA 1292  
 DB 274 aLysProTyrThrSerValAlaLeuAsnArgIleSerArgHisPheArgCysLeuArgAs 294  
 QY 1293 GAGATCTGCAAGAAGATA---ATCTCTATGGATCTGTATTGAGAGAGGCGCAAGACAA 1349  
 DB 294 pAlaIleLysGlnIleGlnValIleArgGlyLysLeuGlyGluArgGluThrSerAs 314  
 QY 1350 GACTCAAGAAACCTCTATGTTCCACCAGCATTCCTTCTTTCAGCAGCTGAAACGAAAGAA 1409  
 DB 314 pGluGlnGlyGluArgIleProArgLeuArgTyrLeuAspGlnArgLeuArgGlnArg 334  
 QY 1410 C-----CATCAG-----ATTGGAGACCTCAAGAGGTTTGCC 1442  
 DB 334 gAlaLeuHisGlnGlnLeuGlyMetValArgProAlaIleArgProGlnArgGlyLeuPr 354  
 QY 1443 TGAGAAATCTGTTTGGTCTTACCGAATGCGATGTTCCAAAACCTCTCTTCAACCTTACCC 1502  
 DB 354 oGluAsnSerValSerIleLeuArgAlaTrpLeuPheGluHisPheLeuHisProTyrPr 374  
 QY 1503 GAAAGATTCGAGAAACATCTTCTAGCTATACGAAGTGGCTTACGAAGAGTCAAGTATC 1562  
 DB 374 oLysGluSerGluLysIleMetLeuSerLysGlnThrGlyLeuSerLysAsnGlnValAl 394  
 QY 1563 AACTGTTTATAAATCGCGGTTAGGCTATGAGCCGATGATGAGAGAGATGATGTC 1622  
 DB 394 aAsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetIleGluMetTyrLy 414  
 QY 1623 GGAA-----ATGAACAAGAGGAGGCTCAATAACAGTTCACATTCACCCCAAGGACC 1673





```

Db      263 rHisGlnMetGluAlaLeuAlaSerSerPheGluMetValThrGlyLeuGlyAlaA 283
QY      1233 ACAGTTTACACCCGGTTGGCTCCCAACCGTTTCTCTTATACAAGAACCTTGAGAGA 1292
Db      283 alyProTyThrSerValAlaLeuAsnArgIleSerArgHisPheArgCysLeuArgAs 303
QY      1293 GAGATCTCCAGAGATA---ATCTTATGGGATCTGTATTGGAGAGAGCAAGACAA 1349
Db      303 palalelyGluGlnIleGlnValleArgGlyLysLeuGlyGluArgGluThrSerAs 323
QY      1350 GACTCAAGAAACCTCTATGTTCCACCAGCATTCCTTCAGAGCTGAACGAAGAA 1409
Db      323 pLugInGlyGluArgIleProArgLeuArgTyLeuAspGlnArgLeuArgGlnArg 343
QY      1410 C-----CATCAG-----ATTGGAGACTCAACGAGTTTGC 1442
Db      343 galAlaLeuHisGlnGlnLeuGlyWetValArgProAlaTrpArgProGlnArgGlyLeuPr 363
QY      1443 TCAGAAATCTGTTTCGGTTCTACGGAATGGATGTCCTCAAACTCTCCACCTTACCC 1502
Db      363 oGluAsnSerValSerIleLeuArgAlaTrpLeuPheGluHisPheLeuHisProTyPr 383
QY      1503 GAAAGATTCGGAGAACATCTTCTAGCTATACGAGTGGCTTGACAAAGTCAGGTATC 1562
Db      383 olysGluSerGluLysIleMetLeuSerLysGlnThrGlyLeuSerLysAsnGlnValAl 403
QY      1563 AAAGTGTTTAAATGCGCGGTTAGGCTATGAGCGCGATGATAGAGAGATGTATGC 1622
Db      403 asenTrpPheIleAsnAlaArgValArgLeuTrpLysProMetLeuGluMetTyLy 423
QY      1623 GGAA-----ATGACACAGAGGAAGTCGTAATACAGTGCACATCAACCAACGGACC 1673
Db      423 sGluGluPheGlyGluSerAlaGluLeuSerAsnSer-----AsnGlnAs 439
QY      1674 AACTCTTGATGCAAAATCTGTTATGATGAGCGCAACCAATGATATAA----- 1722
Db      439 pThrLysLysMetGlnGluThr-----SerGlnLeuLysHisGluAspSerSerSe 456
QY      1723 -----TAAGACACAAATGTGTTTACCAACTTGTGTGATATA 1760
Db      456 rSerGlnGlnAsnGlnGlyAsnAsnAsnAsnIleProTyThrSerAsp-Alag 476
QY      1761 GCCAATGCTACTCTATGATGTCGCCAAACCAACCTTAAACCATGTACCACT 1807
Db      476 luGlnAsnLeuValPheAlaAspProLysProAspArgAlaThrThr 491

RESULT 9
AAE02510
ID   AAE02510 standard; protein; 739 AA.
XX
AC   AAE02510;
DT
DT   10-AUG-2001 (first entry)
DE   Arabidopsis thaliana transcription factor G418.
XX
KW   Transcription factor; pesticidal; antimicrobial; gene therapy;
KW   pathogen tolerance; trichome structure; callose induction;
KW   phytoalexin induction; plant structure; plant development.
XX
OS   Arabidopsis thaliana.
XX
FH   Key
FT   Region
FT   /note= "Conserved domain"
XX
PN   WO200135726-A1.
XX
PD   25-MAY-2001.
XX
PF   14-NOV-2000; 2000WO-US031418.
XX

```

```

PR      17-NOV-1999; 99US-0166228P.
PR      17-APR-2000; 2000US-0197899P.
PR      22-AUG-2000; 2000US-0227439P.
XX
XX      (MEND-) MENDEL BIOTECHNOLOGY INC.
XX      (HEAR/) HEARD J.
XX      (RATC/) RATCLIFFE O.
XX      (CREE/) CREELMAN R.
XX      (JIAN/) JIANG C.
XX      (PINE/) PINEDA O.
XX      (REUB/) REUBER L.
XX      (ADAM/) ADAM L.
XX
PI      Heard J, Ratcliffe O, Creelman R, Jiang C, Pineda O, Reuber L;
PI      Adam L;
DR      WPI; 2001-335978/35.
DR      N-PSDB; AAD06479.
XX
XX      Nucleic acids encoding plant transcription factor polypeptides, useful
XX      for altering the pathogen resistance characteristics of plants, e.g.
XX      corn, potato and cotton plants.
XX
PS      Claim 4; Page 78-81; 134pp; English.
XX
XX      The present sequence is Arabidopsis thaliana transcription factor. The
XX      transcription factors are used to alter the structure and developmental
XX      characteristics of plants such as soybean, wheat, corn, potato, cotton,
XX      rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana,
XX      blackberry, blueberry, strawberry, raspberry, carrot, cantaloupe,
XX      cauliflower, coffee, cucumber, eggplant, grapes, mango, lettuce,
XX      honeydew, melon, onion, papaya, peas, peppers, pineapple, spinach,
XX      squash, sweet corn, tobacco, tomato, watermelon, roseaceous fruits and
XX      vegetable brassicas. The transcription factors are specifically useful
XX      for modifying traits associated with plant's pathogen tolerance such as
XX      alterations in cell wall composition, trichome number or structure,
XX      callose induction, phytoalexin induction, and alterations in the cell
XX      death response. Transgenic plants expressing these transcription factors
XX      are more tolerant to biotrophic or necrotrophic pathogens such as fungi,
XX      bacteria, molluscs, viruses, nematodes and parasitic higher plants. The
XX      transcription factors are also used in gene therapy
XX
SQ      Sequence 739 AA;
Alignment Scores:
Pred. No.: 1,21e-34 Length: 739
Score: 424.00 Matches: 155
Percent Similarity: 41.02% Conservative: 94
Best Local Similarity: 25.54% Mismatches: 171
Query Match: 12.91% Indels: 187
DB: 4 Gaps: 24
US-09-423-575-1 (1-1886) x AAE02510 (1-739)
QY      289 TTTTCATAGAACCCATCGACACAAAC----- 315
Db      17 PheHisAsnSerMetSerGlnAspTyHisHisHisHisHisHisHisGln 36
QY      316 -----AACAAACAACACACTTTTAGTCTCTGGAT-----AATGTCATGACT 357
Db      37 GlyGlyIlePheAsnPheSerAsnGlyPheAspArgSerAspSerProAsnLeuThr 56
QY      358 AACCAAAATCCT-----CTTCTCATGGAT----- 381
Db      57 GlnGlnLysGlnGlnHisGlnArgValGluMetAspGluSerValalaGlyGly 76
QY      382 TTTTATACCTTCAAGAGAGATTCAACTTCATTCCTCAACAACCTTCCTCCATGAATACCATC 441
Db      77 ArgIleProValTyGluSerAlaGlyMetLeuSerGluMetPheAsnPhe----- 93
QY      442 AGATCAGATCTCTCAAAATGGGTGGC-----TTTGATATTTCAATCTATGCTG 492
Db      94 -----ProGlySerSerGlyGlyArgAspLeuAspLeuGlyGlnSerPheArg 110

```





XX Disclosure; SEQ ID NO 236; 470pp; English.

PS The invention relates to a number of cDNA sequence and their encoded

CC proteins which are especially transcription factor cDNAs and their

CC proteins. The isolated or recombinant polynucleotide is useful for

CC producing a modified plant with a modified trait, e.g. enhanced tolerance

CC to environmental conditions, improved tolerance to microbial, fungal or

CC viral diseases, improved tolerance to pest infestation, decreased

CC herbicide sensitivity, improved tolerance of heavy metals, or enhanced

CC ability to take up heavy metals, improved growth under poor

CC photoconditions, improved nutrient uptake, or reduced hormone

CC sensitivity. The transgenic plants are useful for growing a progeny plant

CC comprising the desired trait. The polynucleotides and polypeptides are

CC also useful in bioinformatic search methods. This sequence represents one

CC of the proteins of the invention.

XX SQ Sequence 739 AA;

# Alignment Scores:

Pred. No.: 1,21e-34 Length: 739

Score: 424.00 Matches: 155

Percent Similarity: 41.02% Conservative: 94

Best Local Similarity: 25.54% Mismatches: 171

Query Match: 12.91% Indels: 187

DB: 7 Gaps: 24

US-09-423-575-1 (1-1886) x ADE37297 (1-739)

QY	289	TTTCTAGAACCCAAATGGACAAAC	-----	315
Db	17	PheHisAsnAsnSerMetSerGlnAspTyrHisHisHisHisHisGln		36
QY	316	-----AACCAACAAACACTTTTAGTCTCTGGAT-----	AATGTCATGACT	357
Db	37	GlyGlyIlePheAsnPheSerAsnGlyPheAspArgSerAspSerProAsnLeuThr		56
QY	358	AACCAAAATCCT-----CTTCTCATGAT-----	-----	381
Db	57	GlnGlnLysGlnHisGlnArgValGluMetAspGluGluSerValAlaGlyGly		76
QY	382	TTTATACCTTCAGAGAGATCACTTCATCTCAACAATGCTTCCATGGAATACCATC		441
Db	77	ArgIleProValTyrGluSerAlaGlyMetLeuSerGluMetPheAsn		93
QY	442	AGATCAGATCCTCTACAAATGGTGGC-----TTTGATATTTTCAATTTCTATGCTG		492
Db	94	-----ProGlySerSerGlyGlyArgAspLeuAspLeuGlyGlnSerPheArg		110
QY	493	ACTAACAAATATCTATCATCTTCCACGGTCTATCGATGTTCAAGATAACCGCAATGTT		552
Db	111	SerAsnArgGlnLeuLeuGlu-----GluGlnHisGlnAsnIle		123
QY	553	CAGTTTCATG-----CTTCAT-----	-----	561
Db	124	ProAlaMetAsnAlaThrAspSerAlaThrAlaAlaAlaMetGlnLeuPheLeu		143
QY	562	-----GTCCTCTCTCTCATCTCTCTCCA-----	-----	585
Db	144	MetAsnProProProGlnGlnProProSerProSerSerThrThrSerProArgSer		163
QY	586	-----CTTCAT-----	-----	612
Db	164	HisHisAsnSerSerThrLeuHisMetLeuLeuProSerProSerThrThrHis		183
QY	613	TATGATGATTCCTCAACAAATG-----	-----	636
Db	184	HisGlnAsnTyrThrAsnHisMetSerMetHisGlnLeuProHisGlnHisGlnGln		203
QY	637	-----TGGGGTTTGAAGCAATAGTAGTTTCAGGCAATTTTCAGGTGTAGTTGGT		687
Db	204	IleSerThrTrpGlnSerSerProAspHisHisHisHisHisAsn-----		219

QY	688	CCAAGTGAACCAATGATGCTCTACATTCGGTGAAGAGATTTCCTCGTTCTTAATTTTGGAT		747
Db	220	---SerGlnThrGluIleGlyThrValHisValGluAsn-----SerGlyGly		234
QY	748	AAAAGAAACAATGAGCTTTTCATGAGTCTTGATCAGATGTT-----TCT		792
Db	235	HisGlyGlyGlnGlyLeuSerLeuSerSerSerLeuGluAlaAlaAlaLysAla		254
QY	793	GATGAATGCTCGAGATAAGTCTTTGTGAGCTACAGATTAGCTCA-----		840
Db	255	GluGluTyrArgAsnIleTyrTyrGlyAlaAsnSerSerAsnAlaSerProHisGln		274
QY	841	-----GAGCAAGCTTCTTCAGCAGCAAAAGACATTTCTAATACCTGTGTACT		888
Db	275	TyrAsnGlnPheLysThrLeuLeuAlaAsnSerSerGlnHisHisGlnValLeuAsn		294
QY	889	CAAGGTTTCTCT-----CAACTTATATT		912
Db	295	GlnPheArgSerSerProAlaAlaSerSerSerMetAlaAlaValAlaAsnIleLeuArg		314
QY	913	GGCTCAAAATACCTTCACCTCTGTCAAGAAATACTATCTCATTTCCCGCATACTCGCTC		972
Db	315	AsnSerArgTyrThrThrAlaAlaGlnGluLeuGluGluPheCysSer--ValGlyA		334
QY	973	GATTATTCATCTCGAGAACCGAGTCAGAGCTGCTAGTTCAGCCTTTTACTTTCAGCTTT		1032
Db	334	rgGlyPheLeuLysLysAsnLysLeu-GlyAsnSerSerAsnProAsnThrCysGlyGly		353
QY	1033	GAGAATATACCTGATTTCTTCATGCTGATTTCTTAATACTCGAGCGGTTTCGA---		1089
Db	354	AspGly-----GlyGlySerSerProSerSerAlaGlyAlaAsnLys		367
QY	1090	-----TCTACATTTCAAAGGAGAGCATTAGAACAGAAAGAAACCCATCTC		1134
Db	368	GluHisProProLeuSerAlaSerAspArgIleGluHisGlnArgArgLysValLysLeu		387
QY	1135	TTGGATCTTTCAAATGGTGTGATCGATATAGTATCATTCGCTAGATGAGATTCATAG		1194
Db	388	LeuThrMetLeuGluGluValAspArgArgTyrAsnHisTyrCysGluGlnMetGlnMet		407
QY	1195	GTTATATCAGCGTTC-----CATGTCGACCCGAGCTTAGATCCACAGTTA		1239
Db	408	ValValAsnSerPheAspIleValMetGlyHisGlyAlaAla-----LeuPro		423
QY	1240	CACACCGGTTTGCCTCCAAACCGTTTCTTCTTATCAAGAACCTGAGAGAGAAATC		1299
Db	424	TyrThrAlaLeuAlaGlnLysAlaMetSerArgHisPheArgCysLeuLysAspAlaVal		443
QY	1300	TGCAAGAAGATA-----ATCTCT		1317
Db	444	AlaAlaGlnLeuLysGlnSerCysGluLeuLeuGlyAspLysAspAlaAlaGlyIleSer		463
QY	1318	ATGGATCTGTATTGGAGAGAGGC-----AAAGACAAGACT		1353
Db	464	---SerSerGlyLeuThrLysGlyGluThrProArgLeuArgLeuGluGlnSerLeu		482
QY	1354	CAAGAAACCTCTATGTTCCACCAGCATTCCTTCTTCCAGCAGCTGAAACGAAGAACAT		1413
Db	483	ArgGlnArgAlaPheHisGlnMetGlyMetMetGluGln-----		496
QY	1414	CAGATTGGAGACCTCAACGAGTTTGGCTGAGAAATCTGTTTCGGTTCTACGGAATGG		1473
Db	497	GluAlaTrpArgProGlnArgGlyLeuProGluArgSerValAsnIleLeuArgAlaTrp		516
QY	1474	ATGTTCCAAAACCTTCTTCCACCTTACCGAAAGATTGGGAAACATCTTCTAGCTATA		1533
Db	517	LeuPheGluHisPheLeuHisProTyrProSerAspAlaAspLysHisLeuLeuAlaArg		536
QY	1534	CGAAGTGGCTTCAACAAGAGTCAAGTATCAAACTGGTTTATAAATGCGCGGTTAGGCTA		1593
Db	537	GlnThrGlyLeuSerArgAsnGlnValSerAsnTrpPheIleAsnAlaArgValArgLeu		556
QY	1594	TGGAACCGCATGATAGAGAGATGTATCGGAAATGAACAAGAGGAAGCTCAATACAGT		1653

```
Db      557  TriPlysProMetValGluMetTyfGlnGlnGlnSerLysGluArgGluGlu 576
QY      1654  CACATTCACCAAC 1668
Db      577  GluLeuGluGluAsn 581
RESULT 11
AAG14140
ID  AAG14140 standard; protein; 459 AA.
AC      XX
AC      XX
AC      XX
DT      17-OCT-2000 (first entry)
DE      Arabidopsis thaliana protein fragment SEQ ID NO: 13890.
XX      XX
KW      Protein identification; signal transduction pathway; metabolic pathway;
KW      hybridisation assay; genetic mapping; gene expression control; promoter;
KW      termination sequence.
XX      XX
OS      Arabidopsis thaliana.
XX      XX
PN      EP1033405-A2.
PD      XX
PD      06-SEP-2000.
PF      XX
PF      25-FEB-2000; 2000EP-00301439.
XX      XX
XX      25-FEB-1999; 99US-0121825P.
XX      05-MAR-1999; 99US-0123180P.
XX      09-MAR-1999; 99US-0123548P.
XX      23-NAR-1999; 99US-0125788P.
XX      25-MAR-1999; 99US-0126264P.
XX      29-MAR-1999; 99US-0126785P.
XX      01-APR-1999; 99US-0127462P.
XX      06-APR-1999; 99US-0128234P.
XX      08-APR-1999; 99US-0128714P.
XX      16-APR-1999; 99US-0129845P.
XX      19-APR-1999; 99US-0130077P.
XX      21-APR-1999; 99US-0130449P.
XX      23-APR-1999; 99US-0130510P.
XX      23-APR-1999; 99US-0130891P.
XX      28-APR-1999; 99US-0132449P.
XX      30-APR-1999; 99US-0132048P.
XX      30-APR-1999; 99US-0132407P.
XX      04-MAY-1999; 99US-0132484P.
XX      05-MAY-1999; 99US-0132485P.
XX      06-MAY-1999; 99US-0132486P.
XX      07-MAY-1999; 99US-0132487P.
XX      11-MAY-1999; 99US-0132863P.
XX      14-MAY-1999; 99US-0134256P.
XX      14-MAY-1999; 99US-0134218P.
XX      14-MAY-1999; 99US-0134219P.
XX      14-MAY-1999; 99US-0134221P.
XX      18-MAY-1999; 99US-0134370P.
XX      18-MAY-1999; 99US-0134768P.
XX      19-MAY-1999; 99US-0134941P.
XX      20-MAY-1999; 99US-0135124P.
XX      21-MAY-1999; 99US-0135353P.
XX      24-MAY-1999; 99US-0135629P.
XX      25-MAY-1999; 99US-0136021P.
XX      27-MAY-1999; 99US-0136392P.
XX      28-MAY-1999; 99US-0136782P.
XX      01-JUN-1999; 99US-0137222P.
XX      03-JUN-1999; 99US-0137528P.
XX      04-JUN-1999; 99US-0137502P.
XX      07-JUN-1999; 99US-0137724P.
XX      08-JUN-1999; 99US-0138094P.
XX      10-JUN-1999; 99US-0138540P.
XX      10-JUN-1999; 99US-0138847P.
XX      14-JUN-1999; 99US-0139119P.
XX      16-JUN-1999; 99US-0139452P.
XX      16-JUN-1999; 99US-0139453P.
XX      17-JUN-1999; 99US-0139492P.
XX      18-JUN-1999; 99US-0139454P.
XX      18-JUN-1999; 99US-0139455P.
XX      18-JUN-1999; 99US-0139456P.
XX      18-JUN-1999; 99US-0139457P.
XX      18-JUN-1999; 99US-0139458P.
XX      18-JUN-1999; 99US-0139459P.
XX      18-JUN-1999; 99US-0139460P.
XX      18-JUN-1999; 99US-0139461P.
XX      18-JUN-1999; 99US-0139462P.
XX      18-JUN-1999; 99US-0139463P.
XX      18-JUN-1999; 99US-0139750P.
XX      18-JUN-1999; 99US-0139763P.
XX      21-JUN-1999; 99US-0139817P.
XX      22-JUN-1999; 99US-0139899P.
XX      23-JUN-1999; 99US-0140353P.
XX      23-JUN-1999; 99US-0140354P.
XX      24-JUN-1999; 99US-0140895P.
XX      28-JUN-1999; 99US-0140823P.
XX      29-JUN-1999; 99US-0140991P.
XX      30-JUN-1999; 99US-0141287P.
XX      01-JUL-1999; 99US-0141842P.
XX      02-JUL-1999; 99US-0142154P.
XX      06-JUL-1999; 99US-0142055P.
XX      08-JUL-1999; 99US-0142390P.
XX      09-JUL-1999; 99US-0142803P.
XX      12-JUL-1999; 99US-0142977P.
XX      13-JUL-1999; 99US-0143542P.
XX      14-JUL-1999; 99US-0143624P.
XX      15-JUL-1999; 99US-0144005P.
XX      16-JUL-1999; 99US-0144085P.
XX      16-JUL-1999; 99US-0144086P.
XX      19-JUL-1999; 99US-0144325P.
XX      19-JUL-1999; 99US-0144331P.
XX      19-JUL-1999; 99US-0144332P.
XX      19-JUL-1999; 99US-0144333P.
XX      19-JUL-1999; 99US-0144334P.
XX      19-JUL-1999; 99US-0144335P.
XX      20-JUL-1999; 99US-0144352P.
XX      20-JUL-1999; 99US-0144632P.
XX      20-JUL-1999; 99US-0144884P.
XX      21-JUL-1999; 99US-0144814P.
XX      21-JUL-1999; 99US-0145086P.
XX      21-JUL-1999; 99US-0145088P.
XX      22-JUL-1999; 99US-0145085P.
XX      22-JUL-1999; 99US-0145087P.
XX      22-JUL-1999; 99US-0145089P.
XX      23-JUL-1999; 99US-0145132P.
XX      23-JUL-1999; 99US-0145145P.
XX      23-JUL-1999; 99US-0145218P.
XX      23-JUL-1999; 99US-0145224P.
XX      26-JUL-1999; 99US-0145276P.
XX      27-JUL-1999; 99US-0145913P.
XX      27-JUL-1999; 99US-0145918P.
XX      27-JUL-1999; 99US-0145919P.
XX      28-JUL-1999; 99US-0145951P.
XX      02-AUG-1999; 99US-0146386P.
XX      02-AUG-1999; 99US-0146388P.
XX      02-AUG-1999; 99US-0146389P.
XX      03-AUG-1999; 99US-0147038P.
XX      04-AUG-1999; 99US-0147204P.
XX      05-AUG-1999; 99US-0147302P.
XX      05-AUG-1999; 99US-0147192P.
XX      05-AUG-1999; 99US-0147260P.
XX      06-AUG-1999; 99US-0147303P.
XX      06-AUG-1999; 99US-0147416P.
XX      09-AUG-1999; 99US-0147493P.
XX      09-AUG-1999; 99US-0147935P.
XX      10-AUG-1999; 99US-0148171P.
XX      11-AUG-1999; 99US-0148319P.
XX      12-AUG-1999; 99US-0148341P.
```

PR	13-AUG-1999;	99US-0148565P.	QY	619	GATTCTCTCAACACATGTGGGGTTTTGAAGCAATAGTGAAGTTTCAGGCAITTTTCAGGT	678
PR	13-AUG-1999;	99US-0148684P.	Db	14	AspSerThrAsnAsnVal-----SerProSerLysGlnLeuValLeuSerSer	30
PR	17-AUG-1999;	99US-0149175P.	QY	679	GTAGTTGGTCCCAAGTGAACCAATGATG-----TCATCATTCGGTGAAGA	723
PR	18-AUG-1999;	99US-0149426P.	Db	31	LeuGlyGlyValSerGlnMetValGluIleGlnAspSerGlySerTrpArgAspGlnGlu	50
PR	20-AUG-1999;	99US-0149722P.	QY	724	GATTTCGGGTTTCTAATTTGGAATAAAGAAACAAT-----GAGCTTTTCATTGAGT	774
PR	20-AUG-1999;	99US-0149829P.	Db	51	Asp-----AsnAspA-GasnArgPheProValMetArgLeuGly	64
PR	23-AUG-1999;	99US-0149930P.	QY	775	CTTGCAATCAGATGTTTCTGATGAATGCTCGGAGATAAGTCTTTGTGCAGCTCAAGATT	834
PR	25-AUG-1999;	99US-0150566P.	Db	65	LeuSerSerGlnIleGlu-----	70
PR	26-AUG-1999;	99US-0150884P.	QY	835	GCCTCAGAGCAAGCTTCTTTCGACGACAAAGACATTTCTAATAAC-----GTT	882
PR	27-AUG-1999;	99US-0151066P.	Db	71	-----ThrSerArgGlyAsnAsnAsnAsnGluTrpAlaThrGln	83
PR	27-AUG-1999;	99US-0151080P.	QY	883	GTTACTCAAGGTTTCTCTCAACTATATTGGCTCAAAATACCTTCACTCTCTTCAAGAA	942
PR	30-AUG-1999;	99US-0151303P.	Db	84	ValValSerGlyPheThrArgThrIleHisAsnSerLysTyrLeuLysAlaAlaGlnGlu	103
PR	31-AUG-1999;	99US-0151438P.	QY	943	ATACTATCTCAITTCGCCGCATCTCGCTCGATTATTCATCTCGAGGACCGAGTCAGGA	1002
PR	01-SEP-1999;	99US-0151930P.	Db	104	LeuLeuAspGluThrValAsnVallys-----	112
PR	07-SEP-1999;	99US-0152363P.	QY	1003	GCTGCTAGTTCAGCCTTACTTTCAGCTTTTGAGATATAACTAGTACTTCTTGTGATGAT	1062
PR	07-SEP-1999;	99US-0153070P.	Db	113	LysAlaLeuLysGlnPheGlnProGluGlyAspLysIleAsnGluValLysGlnLysAsn	132
PR	13-SEP-1999;	99US-0153758P.	QY	1063	TCTAATACTCGAGCGGGTTCGGATCTACATTTCAAGAGAGAGATTAGAACCAAG	1122
PR	15-SEP-1999;	99US-0154018P.	Db	133	LeuGlnThrAsnThrAlaGluIleProGlnAla---GluArgGlnGluLeuGlnSerLys	151
PR	16-SEP-1999;	99US-0154039P.	QY	1123	AAAAACCATCTCTGGATCTTCTCAAAATGTTGGATGATCGATATAGTCAITTCGTAGAT	1182
PR	22-SEP-1999;	99US-0155139P.	Db	152	LeuSerLysLeuLeuSerIleLeuAspGluValAspArgAsnTyrLysGlnTyrHis	171
PR	23-SEP-1999;	99US-0155486P.	QY	1183	GAGATCATACGTTATATACGCTTCATGCTCGACCGAGTTAGAT---CCACAGTTA	1239
PR	24-SEP-1999;	99US-0155659P.	Db	172	GlnMetGlnIleValValSerSerPheAspValIleAlaGlyCysGlyAlaAlaLysPro	191
PR	28-SEP-1999;	99US-0156458P.	QY	1240	CACACCGGTTTGGCTCCAAACCGTTTCTTCTTATACAAAGACCTCAGAGAGAGAATC	1299
PR	29-SEP-1999;	99US-0156598P.	Db	192	TyrThrAlaLeuAlaLeuGlnThrIleSerArgHisPheArgCysLeuArgAspAlaIle	211
PR	04-OCT-1999;	99US-0157117P.	QY	1300	TGCAAGAAGATAATC-----TCT	1317
PR	05-OCT-1999;	99US-0157533P.	Db	212	SerGlyGlnIleLeuValIleArgLysSerLeuGlyGluGlnAspGlySerAspGly	231
PR	06-OCT-1999;	99US-0157865P.	QY	1318	ATGGGATCTGTATTGAGAGAGGCAACACAAAGACTCAAGAACTCTATGTTCCACAG	1377
PR	07-OCT-1999;	99US-0158029P.	Db	232	ArgGlyValGlyIleSerArgLeuArgAsnValAspGlnGlnVal-----ArgGln	248
PR	08-OCT-1999;	99US-0158232P.	QY	1378	CATTGCCTTCTTCAGCAGCTGAAGCAAGAACCATCATGATTTCGAGACCTCAACAGGT	1437
PR	12-OCT-1999;	99US-0158366P.	Db	249	GlnArgAlaLeuGlnArgLeuGlyValMetGlnProHisThrTrpArgProGlnArgGly	268
PR	13-OCT-1999;	99US-0158293P.	QY	1438	TTGCCTGAGAATCTGTTTCGGTCTACGGAATGGATGTTCCAAACCTTCTTCCACCT	1497
PR	14-OCT-1999;	99US-0159331P.	Db	269	LeuProAspSerSerValLeuValLeuArgAlaTrpLeuPheGluHisPheLeuHisPro	288
PR	14-OCT-1999;	99US-0159637P.	QY	1498	TACCCGAAAGATTTCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAAGATCAG	1557
PR	18-OCT-1999;	99US-0159584P.	Db	289	TyrProLysAspSerAspLysIleMetLeuAlaArgGlnThrGlyLeuSerArgGlyGln	308
PR	21-OCT-1999;	99US-0160741P.	QY	1558	GTATCAAACTGGTTTATAAATCGCGGGTTAGGCTATGGAAGCCGATGATAGAGAGATG	1617
PR	21-OCT-1999;	99US-0160767P.	Db	309	ValSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetValGluGluMet	328
PR	21-OCT-1999;	99US-0160770P.				
PR	21-OCT-1999;	99US-0160814P.				
PR	21-OCT-1999;	99US-0160815P.				
PR	22-OCT-1999;	99US-0160980P.				
PR	22-OCT-1999;	99US-0160981P.				
PR	22-OCT-1999;	99US-0160989P.				
PR	25-OCT-1999;	99US-0161404P.				
PR	25-OCT-1999;	99US-0161405P.				
PR	25-OCT-1999;	99US-0161406P.				
PR	26-OCT-1999;	99US-0161359P.				
PR	26-OCT-1999;	99US-0161360P.				
PR	26-OCT-1999;	99US-0161361P.				
PR	28-OCT-1999;	99US-0161920P.				
PR	28-OCT-1999;	99US-0161992P.				
PR	28-OCT-1999;	99US-0161993P.				
PR	29-OCT-1999;	99US-0162142P.				

Alignment Scores:  
Pred. No.: 3,25e-34  
Score: 419.00  
Percent Similarity: 47.11%  
Best Local Similarity: 32.51%  
Query Match: 12.76%  
DB: 3  
Length: 459  
Matches: 118  
Conservative: 53  
Mismatch: 120  
Indels: 72  
Gaps: 11  
US-09-423-575-1 (1-1886) x AAG14140 (1-459)

QY 1618 TATGGGAA 1626  
Db ||| |||  
329 TyrLysGlu 331

RESULT 12  
AAG32446  
ID AAG32446 standard; protein; 459 AA.  
XX AC AAG32446;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 39142.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129645P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 03-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140654P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145244P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.



RESULT 13  
AAG32445  
ID AAG32445 standard; protein; 482 AA.  
XX AC AAG32445;  
XX DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 39141.  
XX DE Arabidopsis thaliana.  
XX DE Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0121180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0128845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144866P.  
PR 16-JUL-1999; 99US-0144867P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147044P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147453P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.







```

PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158023P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159323P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 3 32e-34
Score: 419.00
Percent Similarity: 47.11%
Best Local Similarity: 32.51%
Query Match: 12.76%
DB: 3

US-09-423-575-1 (1-1886) x AAG14139 (1-482)

QY 619 GATTCCTCAACACATGTCGGGTTTGAACCAATAGTAGTTCAGGCATTTTCAGGT 678
Db 37 AspSerThrAsnAsnVal-----SerProSerLysGluIleGlnValLeuSerSer 53
QY 679 GTAGTTGGTCCCAAGTGAACCAATGATG-----TCTACATTCGGGTGAAGAA 723
Db 54 LeuGlyGlyValSerGlnMetValGluIleGlnAspSerGlySerTrpArgAspGlnGlu 73

```

```

QY 724 GATTTCCCGTTTCTAATTTCGAATAAAGAAACAAT-----GAGCTTTTCATTGAGT 774
Db 74 Asp-----AsnAspArgAsnArgPheProValMetArgArgLeuGly 87
QY 775 CTTGCATCAGATGTTTCTGATGAATGCTCGGAGATAAGTCTTTGTGCAGCTCAAGATTA 834
Db 88 LeuSerSerGlnIleGlu----- 93
QY 835 GCCTCAGACCAAGCTCTTCGACGACCAAGACATTTCTTAATAAC-----GTT 882
Db 94 -----ThrSerArgGlyAsnAsnAsnAsnGluTrpAlaThrGln 106
QY 883 GTTACTCAAGGTTTCTCTCAACTATATTGGCTCAAAATACCTCACTCTCTTCAAGAA 942
Db 107 ValValSerGlyPheThrArgThrIleHisAsnSerLysTyrLeuLysAlaAlaGlnGlu 126
QY 943 ATACTATCTCATTTCGCCCATCTCGCTCGATATTATTCATCTCGAGGACCGAGTCAGGA 1002
Db 127 LeuLeuAspGluThrValAsnValLys----- 135
QY 1003 GCTGCTAGTTCAGCCTTTACTTCAGTTCGAGATATAACTGAGTTTCTTGATGTTGAT 1062
Db 136 LysAlaLeuLysGlnPheGlnProGluGlyAspLysLysIleAsnGluValLysGluLysAsn 155
QY 1063 TCTAATACTCGGAGCGGGTTTCGATCTACATTTCAAAGAGAGACATTAGAACAAAG 1122
Db 156 LeuGlnThrAsnThrAlaGluIleProGlnAla---GluArgGlnGluLeuGlnSerLys 174
QY 1123 AAAACCCATCTCTGATCTCTTCAATGTTGATGATCGATATAGTCAATTCGCTAGAT 1182
Db 175 LeuSerLysLeuLeuSerIleLeuAspGluValAspArgAsnTyrLysGlnTyrTrpHis 194
QY 1183 GAGATTTCATACGTTTATATCAGCGTTTCCATGCTGCAACCGAGTTAGAT---CCACAGTTA 1239
Db 195 GlnMetGlnIleValValSerSerPheAspValIleAlaGlyCysGlyAlaAlaLysPro 214
QY 1240 CACACCGGTTTGCCTCCAAACCGTTTCTTCTTATACAGAACCTCAGAGAGAGATC 1299
Db 215 TyrThrAlaLeuAlaLeuGlnThrIleSerArgHisPheArgCysLeuArgAspAlaIle 234
QY 1300 TGCAAGAGAGATAATC-----TCT 1317
Db 235 SerGlyGlnIleLeuValIleArgLysSerLeuGlyGlyGluGlnAspGlySerAspGly 254
QY 1318 ATGGGATCTGTTATCGAGAGAGGCAACAGACAGACTCAAGAAACCTCTATGTTCCACCAG 1377
Db 255 ArgGlyValGlyIleSerArgLeuArgAsnValAspGlnGlnVal-----ArgGln 271
QY 1378 CATTCGCTTCTTCAGCAGCTGAACGAAAGAACCATCAGATTTCGAGACCTCAACGAGGT 1437
Db 272 GlnArgAlaLeuGlnArgLeuGlyValMetGlnProHisThrTrpArgProGlnArgGly 291
QY 1438 TTGCTGTGAGAAATCTGTTTCGGTTCACGGAATGAGATGTTCCAAAATCTCTTCACCT 1497
Db 292 LeuProAspSerSerValLeuValLeuArgAlaTrpLeuPheGluHisPheLeuHisPro 311
QY 1498 TACCCGAAAGATTCGAGAAACATCTCTAGCTATACGAAGTGGCTTCACAAGAGTCCAG 1557
Db 312 TyrProLysAspSerAspLysIleMetLeuAlaArgGlnThrGlyLeuSerArgGlyGln 331
QY 1558 GTATCAAACTGGTTTATAAATCGCGGTTAGGCTATCGAAGCCGATGATGAGAGAGATG 1617
Db 332 ValSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetValGluGluMet 351
QY 1618 TATCGGGA 1626
Db 352 TyrLysGlu 354

RESULT 15
AAG21086
ID AAG21086 standard; protein; 491 AA.
XX
AC AAG21086;

```



PR	27-AUG-1999;	99US-0151065P.	
PR	27-AUG-1999;	99US-0151066P.	
PR	27-AUG-1999;	99US-0151080P.	
PR	30-AUG-1999;	99US-0151303P.	
PR	31-AUG-1999;	99US-0151438P.	
PR	01-SEP-1999;	99US-0151930P.	
PR	07-SEP-1999;	99US-0152363P.	
PR	10-SEP-1999;	99US-0153070P.	
PR	13-SEP-1999;	99US-0153758P.	
PR	15-SEP-1999;	99US-0154018P.	
PR	16-SEP-1999;	99US-0154039P.	
PR	20-SEP-1999;	99US-0154779P.	
PR	22-SEP-1999;	99US-0155139P.	
PR	23-SEP-1999;	99US-0155486P.	
PR	24-SEP-1999;	99US-0155659P.	
PR	28-SEP-1999;	99US-0156458P.	
PR	29-SEP-1999;	99US-0156596P.	
PR	04-OCT-1999;	99US-0157117P.	
PR	05-OCT-1999;	99US-0157753P.	
PR	06-OCT-1999;	99US-0157865P.	
PR	07-OCT-1999;	99US-0158029P.	
PR	08-OCT-1999;	99US-0158232P.	
PR	12-OCT-1999;	99US-0158369P.	
PR	13-OCT-1999;	99US-0159293P.	
PR	13-OCT-1999;	99US-0159294P.	
PR	13-OCT-1999;	99US-0159295P.	
PR	14-OCT-1999;	99US-0159329P.	
PR	14-OCT-1999;	99US-0159330P.	
PR	14-OCT-1999;	99US-0159331P.	
PR	14-OCT-1999;	99US-0159637P.	
PR	14-OCT-1999;	99US-0159638P.	
PR	18-OCT-1999;	99US-0159584P.	
PR	21-OCT-1999;	99US-0160741P.	
PR	21-OCT-1999;	99US-0160767P.	
PR	21-OCT-1999;	99US-0160768P.	
PR	21-OCT-1999;	99US-0160770P.	
PR	21-OCT-1999;	99US-0160814P.	
PR	21-OCT-1999;	99US-0160815P.	
PR	22-OCT-1999;	99US-0160980P.	
PR	22-OCT-1999;	99US-0160981P.	
PR	22-OCT-1999;	99US-0160989P.	
PR	25-OCT-1999;	99US-0161404P.	
PR	25-OCT-1999;	99US-0161405P.	
PR	25-OCT-1999;	99US-0161406P.	
PR	25-OCT-1999;	99US-0161359P.	
PR	26-OCT-1999;	99US-0161360P.	
PR	26-OCT-1999;	99US-0161361P.	
PR	28-OCT-1999;	99US-0161920P.	
PR	28-OCT-1999;	99US-0161992P.	
PR	28-OCT-1999;	99US-0161993P.	
PR	29-OCT-1999;	99US-0162142P.	
Alignment Scores:			
Pred. No.:	3.35e-34	Length:	491
Score:	419.00	Matches:	119
Percent Similarity:	48.24%	Conservative:	59
Best Local Similarity:	32.85%	Mismatches:	149
Query Match:	12.76%	Indels:	42
DB:	3	Gaps:	10
US-09-423-575-1 (1-1886) x AAG21086 (1-491)			
QY	746	ATAAAGAAACATGAGCTTTTCATTGCTTGCATCAGATGTTCTCAGTGAATGCTCGG	805
DB	63	VallysGluThrSerProPheHis-----ValAspGluMet--SerVallysSerLysG	80
QY	806	AGATGAAGCTTTGTGTCACGATACAGATAGCTCAGACGACGCTTCTTGACGACGAAAG	865
pb	80	luMetileLeuLeuGlyGlnSerAspProSerGlyTyrAlaGlyAsnGlyGlyAsnG	100
QY	866	ACATTCTTAATAACGTT-----GTTACTCAAGGTTTCTCTCAACTATATTTG	913
DB	100	lyPheTyAsnAsnTyrArgTyrAsnGluThrSerGlyGlyPheMetSerSerValLeuA	120

QY	914	GCTCAAAATACCTTCACTCTGTTCAAGAAATACTATCTTCAATTCGCGCATCTCGCTCG	973
DB	120	rgSerArgTyrLeuLysProAlaGlnAsnLeuLeuAspValValSerValLysG	140
QY	974	ATTATTCATCTCGAAGAACCGAGTCAGGCTGCTAGTTCAGCCTTTACTTCACTTTTG	1033
DB	140	luLeuAsnGlnMetSerLysLysMetLysValAsnAsp-----PheA	155
QY	1034	AGATATACTAGTTCTTGTATGTT-----GATTTCTAATA	1069
DB	155	snAsnGlySerLysGluIleGluGlyGlySerGlyLeuSerAsnAspSerAsnG	175
QY	1070	ACTCGGAGCGGTTTCGGATCTACATTTCAAAGGAGAGCATTAGAAGCAAGAAACCC	1129
DB	175	lyLysSerileGluLeu---SerThrileGluArgAspGluLeuHisAsnLysAsnL	194
QY	1130	ATCTCTTGGATCTTCTCAATGCTGGATGATCGATATAGTCATTCGCTAGATGATTC	1189
DB	194	ysLeuLeuThrMetValAspGluValAspLysArgTyrAsnGlnTyrTyrHisGlnMetG	214
QY	1190	ATACGGTTATATCAGCGTTTCCATGCTGCAACCGAGTTA---GATCCACAGTTTACACCC	1246
DB	214	luAlaLeuAlaSerSerPheGluIleValAlaGlyLeuGlySerAlaLysProTyrThr	234
QY	1247	GTTTGGCCCTCCAAACCGTTTCTTCTTATACAAGAACCTGAGAGAGATCTCAAGA	1306
DB	234	erValAlaLeuAsnArgileSerArgHisPheArgAlaLeuArgAlaileLysGluG	254
QY	1307	AGATA---ATCTCTATGGATCTGATTTGGAGAGAGCAAGACAAAGACTCAAGAACT	1363
DB	254	inileGlnileValargGluLysLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyG	274
QY	1364	CT-----ATGTTCCACGACGATTC	1384
DB	274	lnGlyGluArgileProArgLeuArgTyrLeuAspGlnArgLeuArgGlnArgAlaL	294
QY	1385	TTCTTCAGCAGCTGAACAAAGAACCATCAGATTTGGAGACCTCAACAGAGTTTGCCTG	1444
DB	294	euHisGlnGlnLeuGlyMetVal---ArgProAlaTrpArgProGlnArgGlyLeuProG	313
QY	1445	AGAAATCTGTTTCGGTTCTTACGGAATGGATGTTTCCAAACTTCCTTCCCTTACCCGA	1504
DB	313	luAsnSerValSerValLeuArgAlaTrpLeuPheGluHisPheLeuHisProTyrProL	333
QY	1505	AGATTCGGAGAAACATCTTCTAGCTATACGAGTGGCTTGACAGAGAGTCAAGTATCAA	1564
DB	333	ysGluSerGluLysileMetLeuAlaLysGlnThrGlyLeuSerLysAsnGlnValAlaA	353
QY	1565	ACTGTTTATAAATCGCGGTTAGGCTATGAAGCCGATGATGAGAGAGATGTATGCGG	1624
DB	353	snTrpPheileAsnAlaargValargLeuTrpLysProMetileGluMetLysG	373
QY	1625	AA-----ATGAACAAGAGAGAGCTCAATPACAGTCACATTCACACCAACCGAC	1672
DB	373	luGluPheGlyAspGluSerGluLeuLeuileSerLysSerSerGlnGluProAsnSerT	393
QY	1673	CAACTCTTCGAATGCCAAATCTGTTATGATGAGCAGCAATGATTAATAAGCAACA	1732
DB	393	hrAsnGlnLysSerSerSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG	413
QY	1733	ATTGTTTACCACCTTTGTGAT	1755
DB	413	snLeuAlaTyrSerSerAlaAsp	420

Search completed: September 2, 2004, 12:06:52  
Job time : 140.5 secs

Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 202, App  
Sequence 2, Appli  
Sequence 15, Appl  
Sequence 14, Appl  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 6, Appli  
Sequence 2, Appli  
Sequence 478, App  
Sequence 485, App  
Sequence 478, App  
Sequence 485, App  
Sequence 485, App  
Sequence 4, Appli  
Sequence 6, Appli  
Sequence 10, Appl  
Sequence 14, Appl  
Sequence 9, Appli  
Sequence 13, Appl  
Sequence 4, Appli  
Sequence 25, Appli  
Sequence 9, Appli

13 106.5 3.2 971 3 US-08-918-190-2  
14 106.5 3.2 971 3 US-09-234-232-2  
15 106.5 3.2 971 5 PCT-US96-09927-2  
16 104.5 3.2 1153 4 US-09-363-137-202  
17 104 3.2 596 4 US-09-752-165-2  
18 104 3.2 990 2 US-08-645-193B-15  
19 102.5 3.1 1153 3 US-08-665-574C-14  
20 102.5 3.1 1153 3 US-08-946-994-14  
21 100.5 3.1 1147 3 US-08-470-260-5  
22 100.5 3.1 1147 3 US-08-471-491-5  
23 100.5 3.1 1147 3 US-08-466-662-5  
24 100.5 3.1 1863 1 US-08-798-691-4  
25 100.5 3.1 1863 3 US-08-825-487A-4  
26 100.5 3.1 1863 3 US-09-074-476-6  
27 100.5 3.1 3289 2 US-08-477-451-2  
28 100 3.0 505 4 US-09-620-405B-478  
29 100 3.0 505 4 US-09-620-405B-485  
30 100 3.0 505 4 US-09-604-287A-478  
31 100 3.0 505 4 US-09-604-287A-485  
32 100 3.0 505 4 US-09-834-759-478  
33 100 3.0 505 4 US-09-834-759-485  
34 100 3.0 3079 5 PCT-US94-00198-4  
35 99.5 3.0 1056 4 US-09-710-092-6  
36 99.5 3.0 1270 4 US-09-710-092-10  
37 99.5 3.0 1426 4 US-08-801-263A-9  
38 99.5 3.0 2512 2 US-09-102-248-9  
39 99.5 3.0 2512 3 US-09-367-764-9  
40 99 3.0 1153 1 US-08-097-997A-14  
41 98 3.0 3665 2 US-08-222-617A-13  
42 98 3.0 3712 2 US-08-222-617A-4  
43 98 3.0 3712 2 US-08-222-617A-25  
44 98 3.0 3712 2 US-08-222-617A-25  
45 97.5 3.0 1169 1 US-08-620-717A-9

ALIGNMENTS

RESULT 1  
US-09-621-976-4543  
; Sequence 4543, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Joebert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621.976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 4543  
; LENGTH: 142  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-621-976-4543

Alignment Scores:  
Pred. No.: 9.08e-08 Length: 142  
Score: 151.00 Matches: 31  
Percent Similarity: 55.67% Conservative: 23  
Best Local Similarity: 31.96% Mismatches: 31  
Query Match: 4.60% Indels: 12  
DB: 4 Gaps: 2  
US-09-423-575-1 (1-1886) x US-09-621-976-4543 (1-142)  
QY 1324 TCTTATTTGGAGAGGCAAGCAAGCACTTATGTTCCACGACATTGC 1383  
Db 31 SerilleMetSerArgAsnAsnAlaaspThrGlyArgValLeuAlaLeuProGluHis--- 49  
QY 1384 CTCTTTCAGCAGCTGAACAGCAAGCAACATCAGATTGGAGACCTCAACGAGGTTTGCT 1443

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: September 2, 2004, 12:02:38 ; Search time 24.5 seconds  
(without alignments)  
7948.292 Million cell updates/sec

Title: US-09-423-575-1  
Perfect score: 3284  
Sequence: 1 attaggtataaaattgtgc.....ttaaaaaaaaaaaaaaaaaa 1886

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DBV=xlh  
-O=/cgn2\_1/USPTO.spool/US09423575/runat\_01092004.161107.18624/app.query.fasta.1.2055  
-DB=Issued\_Patents\_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPECL=0  
-LOOPEXT=0 -UNITS=5bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09423575 @CGN 1.1 26 @runat\_01092004.161107.18624 -NCPU=6 -ICFU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOB=6  
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/SA\_COMB.pep.\*  
2: /cgn2\_6/prodata/2/iaa/SB\_COMB.pep.\*  
3: /cgn2\_6/prodata/2/iaa/SA\_COMB.pep.\*  
4: /cgn2\_6/prodata/2/iaa/SB\_COMB.pep.\*  
5: /cgn2\_6/prodata/2/iaa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	151	4.6	142	4	US-09-621-976-4543
2	124.5	3.8	2954	4	Sequence 4543, Ap
3	115	3.5	1186	2	Sequence 1, Appli
4	115	3.5	1186	2	Sequence 8, Appli
5	115	3.5	1186	3	Sequence 8, Appli
6	114.5	3.5	676	4	Sequence 4318, Ap
7	112.5	3.4	3969	3	Sequence 5, Appli
8	110	3.3	944	4	Sequence 27, Appli
9	109.5	3.3	1584	4	Sequence 3678, Ap
10	108	3.3	522	4	Sequence 5, Appli
11	107	3.3	512	2	Sequence 2, Appli
12	106.5	3.2	971	1	Sequence 2, Appli

Db 50 -----LysLysLysArgLysGlyAsn-----LeuPro 58  
 QY 1444 GAGAAATCTGTTTCGGTTCTACGAATTCGATGTTCCAAATCTCTTACCCG 1503  
 Db 59 AlaGluSerValLysIleLeuArgAspTrpMetTyrLysHisArgPheLysAlaTyrPro 78  
 QY 1504 AAAGATTCCGAGAAACATCTTAGCTATACGAAGTGGGTTGACAGAGTCAAGATCA 1563  
 Db 79 SerGluGluGlnMetLysGlnMetLysGlnMetLysGlnMetLysGlnMetLys 98  
 QY 1564 AACTGGTTTAAATCGCGGGTTAGGCTATGGAAGCCGATGATAGAAGAG 1614  
 Db 99 AsnTrpPheIleAsnAlaArgArgGileLeuProAspMetLeuGlnGln 115

RESULT 2  
 US-09-150-867-1  
 / Sequence 1, Application US/09150867  
 / Patent No. 6645748  
 / GENERAL INFORMATION:  
 / APPLICANT: Wood, Kenneth W.  
 / APPLICANT: Sakowicz, Roman  
 / APPLICANT: Goldstein, Lawrence S.B.  
 / APPLICANT: Cleveland, Don W.  
 / APPLICANT: The Regents of the University of California  
 / TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for  
 / TITLE OF INVENTION: Chromosome Congression  
 / FILE REFERENCE: 18557C-000110US  
 / CURRENT APPLICATION NUMBER: US/09/150,867  
 / EARLIER FILING DATE: 1998-09-10  
 / EARLIER FILING DATE: US 60/058,645  
 / EARLIER FILING DATE: 1997-09-11  
 / NUMBER OF SEQ ID NOS: 11  
 / SOFTWARE: PatentIn Ver. 2.0  
 / SEQ ID NO 1  
 / LENGTH: 2954  
 / TYPE: PRT  
 / ORGANISM: Xenopus sp.  
 / FEATURE:  
 / OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)  
 / OTHER INFORMATION: member of the kinesin superfamily of microtubule  
 / OTHER INFORMATION: motor proteins  
 / FEATURE:  
 / NAME/KEY: DOMAIN  
 / LOCATION: (1)..(472)  
 / OTHER INFORMATION: kinesin like motor domain  
 / FEATURE:  
 / NAME/KEY: DOMAIN  
 / LOCATION: (473)..(2752)  
 / OTHER INFORMATION: rod domain  
 / FEATURE:  
 / NAME/KEY: DOMAIN  
 / LOCATION: (2753)..(2954)  
 / OTHER INFORMATION: tail domain  
 US-09-150-867-1

Alignment Scores:  
 Pred. No.: 0.000415 Length: 2954  
 Score: 124.50 Matches: 103  
 Percent Similarity: 34.74% Conservative: 78  
 Best Local Similarity: 19.77% Mismatches: 205  
 Query Match: 3.79% Indels: 135  
 DB: 4 Gaps: 21

US-09-423-575-1 (1-1886) x US-09-150-867-1 (1-2954)  
 QY 142 CTTTCACGAGAAAGAGATAAAT-----AAGCATTATCTTCTTCTAGTTTAAACA 195  
 Db 1871 LeuSerGluA:GAspAspLeuAsnGlnSerLysGlnHisLeuPheSerGluGluThr 1890  
 QY 196 CACATTTTGGAAATTTTCATGCTAAATTTCTTTTGGACGTTGTTGTTGTTGTTGTT 255  
 Db 1891 LeuSerLeuSerLeuLysGluLysGluPheAlaLeuGluGlnAla-----GluLys 1907

QY 256 CCCAAGGTTCTATCAGAGAAGAGATAAAGTTTCTATAGAAACCAATGGACAACAAC 315  
 Db 1908 AsplysAlaAspAlaAlaArgLysThrIleAspIle----- 1919  
 QY 316 AACACACACAACTTTTAGTTCTCTGGTAATGTCTAGTACTAACCACAAATCCTCTTCTC 375  
 Db 1920 -----ThrGluLysIleSerAsnIleGluGluGlnLeuGlnGlnAlaThrAsnLeu 1937  
 QY 376 ATGGATTTTATACCTTCAAGAGAAGATTCAACTTCAATCTTCAACAATGTTCCATGAAT 435  
 Db 1938 LysGluThrLeuTyrGluArgGluSerLeuIleGlnCysLysGluGlnLeuAlaLeuAsn 1957  
 QY 436 -----ACCATCAGATCAGATCTCTACAATGGTGGCTTGTGATT 477  
 Db 1958 ThrGluHisLeuArgGluThrLeuLysSerLysAspLeuAlaLeuGlyLysMetGluGln 1977  
 QY 478 TTCAATTCTATGCTGACTAACAAATATTATCATCTTCTCCACGG-----TCTATCGAT 531  
 Db 1978 GluArgAspGluAlaAlaAsnLysValIleAlaLeuThrGluLysMetSerSerLeuGlu 1997  
 QY 532 GTTCAAGATAACCGCAATGTTGAGTTTCATGGCTCTCTCTCTCATCTCTCTCTCTCAT 591  
 Db 1998 GluGlnIleAsnGluAsnValThrThrLeuLys----- 2008  
 QY 592 CTTTGGATCATTTAAGACACTATGATGATTCTCTCAACACACATGTGGGTTTGAAGCA 651  
 Db 2009 -----GluGly 2010  
 QY 652 AATAGTGGTTTCAGGCATTTTCAGGTGTAGTGTGTCCAAGTGAACCAATGATGTCTACA 711  
 Db 2011 GluGlyGluLysGluThrPheTyr---LeuGlnArgProSerLysGlnGlnSerSer 2029  
 QY 712 TTCGGTGAAGAGATTCCCGTTTCTAATTTCGAATAAAGAAACAATGAGCTTTTCATTG 771  
 Db 2030 GlnMetGluGlu-----LeuArgGluSerLeuLysThrLysAspLeuGlnLeu 2045  
 QY 772 AGTCTTTCGCA---TCAGATGTTCTGTATGAATGCTCGAGATAAGTCTTTGTGCAGCTACA 828  
 Db 2046 GluGluAlaGluLysGluIleSerGluAlaThrAsnGluIleLysAsnLeuThrAlaLys 2065  
 QY 829 AGATTAGCCTCAGACCAAGCTTTCTCCAGCACAAAGACATTCTTAATAACCTTTACT 888  
 Db 2066 IleSerSerLeuGluGluGluIleLeuGlnAsnAlaSerIleLeuAsnGluAlaValSer 2085  
 QY 889 CAAGGTTTCTCTCACTTATATTGGCTCAAAATACCTTCACTCTGTTCAAGAAATACTA 948  
 Db 2086 Glu-----ArgGluAsnLeuArgHisSerLysGlnGlnLeuVal 2098  
 QY 949 TCTCATTTCCGCGCATCTCGCTCGATTTATTCATCTCGAGAACCGAGTCAGAGCTGCT 1008  
 Db 2099 SerGluLeuGluGlnLeuSerLeuThrLysSerArg----- 2111  
 QY 1009 AGTTCAGCTTTTACTTCAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTAT 1068  
 Db 2112 AspHisAlaPheAlaGlnSerLysArgGluLysAspGluAlaValAsnLysIleAlaSer 2131  
 QY 1069 AACTCGAGCGGGTTTCGGATCTTACATTTCAAAGAGAGCATTTAGAGCAAGAAACC 1128  
 Db 2132 LeuAlaGlu-----GluIleLysIleLeuThrLysGluMetAsp 2144  
 QY 1129 CATCTCTTGGATCTTCTTCAATGTTGATGATGATGATGATGATGATGATGATGATGAT 1188  
 Db 2145 GluPheArgAspSerLysGluSerLeuGlnGlnSerSerHisLeuSerGluGluLeu 2164  
 QY 1189 CATACGGTTTATACAGCTTCCATCTGCAACCGAGTTAGATCCACAGTTACACACCGG 1248  
 Db 2165 CysThrTyr-----LysThrGluLeuGln----- 2172  
 QY 1249 TTTGCCCTCCAAACCGTTTCTTCTTATACAGAACCTTGAGAGAGAGATCTGCAAGAG 1308  
 Db 2173 -----MetLeuLysGlnGlnLysGluAspIleAsnAsnLys 2184  
 QY 1309 ATAATCTCTATGGATCTGTATTGGAGAGCGCAAGACAAGACTCAAGAAACCTTATG 1368



```

Db      945 IleProSerAspGlnGlnAsnGluMetVal-----ArgGlu 956
QY      1111 TTAGAAGCAAGAAACCCATCTCTGGATCTTCCTCAATGGTGATCATGATATAGT 1170
Db      957 LeuAspGly-----HisValLeuLysCysValLys-----AspGlnAsnGlyAsn 971
QY      1171 CAT-----TGCCTAGAT-----GAGATTCATACGGTGTATATCA 1203
Db      972 HisValValGlnLysCysValGlnProGlnSerLeuGlnPheIleLeuAsp 991
QY      1204 GCGTTCATCGTCACACCGAGTAGATCACAGTTACACACCGGTTGCGCTCCAAACC 1263
Db      992 AlaPheLysGlyGlnVal-----PheAlaLeuSerThr 1002
QY      1264 GTTCTCTCTTATACAAAGAACCTGAGAGAGAGAAATCTGCAAGAGATATCTCTATGGGA 1323
Db      1003 HisProTyr-----GlyCysArgValIleGlnArgIleLeu----- 1014
QY      1324 TCTGTATTGGAGAGCGAAGAACAGACTCAAGAAACCTCTATGTTCCACGAGCATTCG 1383
Db      1015 -----GlnHisCys 1017
QY      1384 CTT-----CTTCAGCAGCTGAAGAACGAAAGAACCATCAGATTGG 1422
Db      1018 LeuProAspGlnThrLeuProIleLeuGluLeuHisGlnHisThrGluGlnLeuVal 1037
QY      1423 AGACCTCAACGAGGTTGCTGCTGAGAAATCTGTTTCGGTTCACGGAATTGGATGCCAA 1482
Db      1038 GlnAspGlnTyrGly-----AsnTyrValIleGln 1047
QY      1483 AACTTCCTT-----CACCTTACCGGAAGATTCCGAGAAACATCTCTACGCTATACGAAGT 1539
Db      1048 HisValLeuGluHisGlyArgProGluAspLysSerLysIleValAlaGluIleArgGly 1067
QY      1540 GGC----- 1542
Db      1068 AsnValLeuValLeuSerGlnHisLysPheAlaSerAsnValValGluLysCysValThr 1087
QY      1542 ----- 1542
Db      1088 HisAlaSerArgThrGluArgAlaValLeuIleAspGluValCysThrMetAsnAspGly 1107
QY      1543 -----TTGCAAGAAAGTCAGGTATCAAACTGATTAATAAATGCG 1591
Db      1108 ProHisSerAlaLeuTyrThrMetMetLysAspGlnTyrAlaAsnTyrValVal----- 1125
QY      1592 CGGGTAGGCTATGAAGCCGATGATAGAGAGATGTATGCGAAATGAACAAAGAGGAAG 1641
Db      1126 -----GlnLysMetIleAsp-----ValAlaGluProGlyGlnArgLys 1138
QY      1642 CTCATTAACAGTCACATTCAACCCAGGACCAACTCTTCGA 1683
Db      1139 IleValMetHisLysIleArgProHisIleAlaThrLeuArg 1152

RESULT 4
US-08-396-001-8
; Sequence 8, Application US/08396001
; Patent No. 5919618
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Claus, James
; APPLICANT: Cole, Francesca
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: Genes Determining Cellular Senescence in
; Yeast
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA

```

```

; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/396,001
; FILING DATE: 28-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6408A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-396-001-8

```

```

Alignment Scores:
Pred. No.: 0.00271 Length: 1186
Score: 115.00 Matches: 108
Percent Similarity: 31.77% Conservative: 68
Best Local Similarity: 19.49% Mismatches: 180
Query Match: 3.50% Indels: 198
DB: 2 Gaps: 25
US-09-423-575-1 (1-1886) x US-08-396-001-8 (1-1186)

```

```

QY      319 AACAAACACACITTTAGTTCTCTGGNTAATGTCATGACTAACCAAAATCCTCTCTCATG 378
Db      698 AsnSerAsnThrGlySerGly----- 704
QY      379 GATTTTATACCTTCAAGAGAGATTCACTTCATTCTCAACAATGCTTCATGGAATACC 438
Db      705 -----SerArgAspSerLeuThrGlySerAspLeuTyrIysArgThr 720
QY      439 ATCAGATCAGATCCTCTACAAATGGTGGCTTGTATATTTCAATCTATGCTGACTAAC 498
Db      721 -----SerSerSerLeuThrProIleGlyHisSerPheTyrAsnGlyLeu----- 735
QY      499 AAATACTTATCATCTCTCCAGGCTATCGATGTTCAAGATAACCGCAATGTCGAGTTC 558
Db      735 ----SerPheSerSerSerProGlyProValGlyMet----- 746
QY      559 ATGGCTCCTCCTCTCATCTCTCCACTTCAAT-----CCTTGGATCATTTA 606
Db      747 -----ProLeuProSerGlnGlyProGlyHisSerGlnThrProProProSerLeuSer 764
QY      607 AGACATCATGATGATTCCTCAACAAACATGCGGGTGTGAGCAAAATAGT----- 657
Db      765 SerHisGlySerSerSerSerLeuAsnLeuGlyGlyLeuThrAsnGlySerGlyArgTyr 784
QY      658 -----GAGTTTCAGGCATTTTTCAGGTGTA 681
Db      785 IleSerAlaAlaProGlyAlaGluAlaLysTyrArgSerAlaSerAlaSerSerLeu 804
QY      682 GTTGCTCAAGTGAACCAATGATGCTACA-----TTCGGTGAAGACATTC 729
Db      805 PheSerProSerSerThrLeuPheSerSerSerArgLeuArgTyrGlyMetSerAspVal 824
QY      730 -----CCGTTTCTAATTTTCGAATAAAGAAACAAATGAGCTTTCATTGAGT 774
Db      825 MetProSerGlyArgSerArgLeuLeuAspPheArgAsnAsnArgTyrProAsnLeu 844
QY      775 CTTGTCATCAGATGTTTCTGTGATGATGTCGCGAGATAAGTCTT-----TGTGCA 822

```



Db 845 GlnLeuArgGluIleAlaGlyHisIleMetGluPheSerGlnAspGlnHisGlySerArg 864  
QY 823 GCTCAAGATTAGCTCTCAGCAAGCTTCTGCAGCAGCAAGACATTTCTAATAAGCTT 882  
Db 865 PheIleGlnLeuLysLeuGluArgAlaThrProAlaGluArgGlnLeuValPheAsnGlu 884  
QY 883 GTTACTCAAGGTTTCTCAACTT-----ATATTGGCTCAAAATACCTTCACCTCT 933  
Db 885 IleLeuGlnAlaAlaTyGlnLeuMetValAspValPheGlyAsnTyValIleGlnLys 904  
QY 934 GTTCAAGAAATCTATCTCATTTCCCGCATCTCGTATTCATCTCTCGAGGAC 993  
Db 905 PhePheGluPheGlySerLeuGlnLysLeuAlaLeuAlaGluArgIleArgGlyHis 924  
QY 994 GAGTCAGGAGCTGCTAGTTTCAGCTTTTACTTCACGT---TTTGAGATATAACTCAGTTT 1050  
Db 925 ValLeuSerLeuAlaLeuGlnMetTyGlyCysArgValIleGlnLysAlaLeuGluPhe 944  
QY 1051 CTTGATGGTATCTTAATAACTCGGAGCGGGTTTCGGATCTACATTTCAAGAGAGCA 1110  
Db 945 IleProSerAspGlnGluAsnGluMetVal-----ArgGlu 956  
QY 1111 TTAGAAGCAAGAAACCATCTCTTGGATCTTCTCAATGGTGGATGATCGATATAGT 1170  
Db 957 LeuAspGly-----HisValLeuLysCysValLys-----AspGlnAsnGlyAsn 971  
QY 1171 CAT-----TGCTAGAT-----GAGATTCATACGGTTTATATCA 1203  
Db 972 HisValValGlnLysCysIleGluCysValGlnProGlnSerLeuGlnPheIleIleAsp 991  
QY 1204 GGGTTCCATCTGCAACCGAGTAGATCCACAGTTACACCGCGTTTCCTCCAAACC 1263  
Db 992 AlaPheLysGlyGlnVal-----PheAlaLeuSerThr 1002  
QY 1264 GTTCTCTCTTATACAGAACCTGAGAGAGAAATCTGCAAGAGATATCTCTATGGGA 1323  
Db 1003 HisProTyR-----GlyCysArgValIleGlnArgIleLeu----- 1014  
QY 1324 TCTGATTGGAGAGGGCAAGACAGACTCAAGAAACCTCTATGTTCCACGAGATTGC 1383  
Db 1015 -----GluHisCys 1017  
QY 1384 CTT-----CTTCAGAGCTGAAACGAAAGAACCATCAGATTTGG 1422  
Db 1018 LeuProAspGlnThrLeuProIleLeuGluLeuHisGlnHisThrGluGlnLeuVal 1037  
QY 1423 AGACCTCAACGAGTTTGCCTAGAAATCTGTTCCGGTTCTACGGAATGGATGTTCCAA 1482  
Db 1038 GlnAspGlnTyGly-----AsnTyValIleGln 1047  
QY 1483 AACTTCCTT---CACCTTACCGAAGATTGGAGAAATCTTCTAGCTATACGAAT 1539  
Db 1048 HisValLeuGluHisGlyArgProGluAspLysSerLysIleValAlaGluIleArgGly 1067  
QY 1540 GGC----- 1542  
Db 1068 AsnValLeuValLeuSerGlnHisLysPheAlaSerAsnValValGluLysCysValThr 1087  
QY 1542 ----- 1542  
Db 1088 HisAlaSerArgThrGluArgAlaValLeuIleAspGluValCysThrMetAsnAspGly 1107  
QY 1543 -----TTGACAAGAAGTCAGGTATCAACTGAGTTTAAATGCG 1581  
Db 1108 ProHisSerAlaLeuTyThrMetMetLysAspGlnTyAlaAsnTyValVal----- 1125  
QY 1582 CGGGTTAGGCTATGGAAGCGCATGATAGAGAGATGATGCGGAATGAAACAGAGGAAG 1641  
Db 1126 -----GlnLysMetIleAsp-----ValAlaGluProGlyGlnArgLys 1138  
QY 1642 CTCAATAACAGTCACATTCAACCAACCGACCACTCTCGA 1683  
Db 1139 IleValMetHisLysIleArgProHisIleAlaThrLeuArg 1152

## RESULT 5

US-09-323-433A-8  
; Sequence 8, Application US/09323433A  
; Patent No. 6218512  
; GENERAL INFORMATION:  
; APPLICANT: Guarente, Leonard P.  
; APPLICANT: Austriaco Jr., Nicanor  
; APPLICANT: Claus, James J.  
; APPLICANT: Cole, Francesca  
; APPLICANT: Kennedy, Brian  
; TITLE OF INVENTION: GENES DETERMINING CELLULAR SENESCENCE IN  
; TITLE OF INVENTION: YEAST  
; FILE REFERENCE: 0050.1491-003  
; CURRENT APPLICATION NUMBER: US/09/323,433A  
; CURRENT FILING DATE: 1999-06-01  
; PRIOR APPLICATION NUMBER: US 08/396,001  
; PRIOR FILING DATE: 1995-02-28  
; PRIOR APPLICATION NUMBER: PCT/US94/09351  
; PRIOR FILING DATE: 1994-08-15  
; PRIOR APPLICATION NUMBER: US 08/107,408  
; PRIOR FILING DATE: 1993-08-16  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 8  
; LENGTH: 1186  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-323-433A-8

Alignment Scores:  
Pred. No.: 0.00271 Length: 1186  
Score: 115.00 Matches: 108  
Percent Similarity: 31.77% Conservative: 68  
Best Local Similarity: 19.49% Mismatches: 180  
Query Match: 3.50% Indels: 198  
DB: 3 Gaps: 25

US-09-423-575-1 (1-1886) x US-09-323-433A-8 (1-1186)

QY 319 AACAAACACTTTTATGTTCTCTGGATAATGTCATGACTAACCAAAATCCTCTCTCATG 378  
Db 698 AsnSerAsnThrGlySerGly----- 704  
QY 379 GATTTTATACCTTCAAGAGAAGATTCAACTTCATTCCTCAACAATGCTTCCATGGAATACC 438  
Db 705 -----SerArgArgAspSerLeuThrGlySerSerAspLeuTyLysArgThr 720  
QY 439 ATCAGATCAGATCCTCTACAAATGGGTGGCTTTGATATTTTCAATTCATCTGCTGACTAAC 498  
Db 721 -----SerSerSerLeuThrProIleGlyHisSerPheTyAsnGlyLeu----- 735  
QY 499 AATACTTATCATCTTCTCCACGGTCTATCGATGTTCAAGATAACCGCAATGTTGAGTTC 558  
Db 736 ---SerPheSerSerProGlyProValGlyMet----- 746  
QY 559 ATGGCTCTCTCTCTCATCTCTCTCCACTTCAT-----CCTTTGGATCATTTA 606  
Db 747 -----ProLeuProSerGlnGlyProGlyHisSerGlnThrProProSerLeuSer 764  
QY 607 AGACACTATGATGTTCTCTCAACAACATGTTGGGTTTGAACAATAGT----- 657  
Db 765 SerHisGlySerSerSerSerLeuAsnLeuGlyLeuThrAsnGlySerGlyArgTyR 784  
QY 658 -----GAGTTTCAGGCATTTTCAGGTGTA 681  
Db 785 IleSerAlaAlaProGlyAlaGluAlaLysTyArgSerAlaSerAlaSerSerLeu 804  
QY 682 GTTGGTCCAATGAAACCAATGATCTCTACA-----TTCGGTGAAGAAGATTTTC 729  
Db 805 PheSerProSerSerThrLeuPheSerSerSerArgLeuArgTyRgGlyMetSerAspVal 824  
QY 730 -----CCGTTTCTAATTCGAATAAAGAAACAATGAGCTTTCATTGAGT 774



```
QY 772 AGTCTGCATCAGATGTT-----TCTGATGAATGCTCGGAG 807
Db 420 ThrSerSerAspAspIleLeuLysSerMetLeuAspGlnSerLysAspLysGluSerLeu 439
QY 808 ATAAGTCTTTGTGAGGTACAGATTAGCTTCAGCAGCAAGCTTCTTCAGCAGCAAGAC 867
Db 440 IleLysGlnLeuLeuThrThrArgLeuGlyAsnAspGluAlaAspArgIleAlaLysLys 459
QY 868 ATTCTTAATACGTTGTACTCAAGGTTCTCTCAACTATATATTGGCTCAAAATACCTT 927
Db 460 Leu-----LeuSerGlnAsnLeuSerAsnSerGlnIle--- 470
QY 928 CACTCTGTTCAAGAAATACTATCTCATTTCCCGCATACTCGCTCGATTATTCTCTCGA 987
Db 471 -----ValGluGlnLeuLysArgHisPhe-----AsnSerGln 481
QY 988 GGAACCGAGTCAGGAGCTGTAGTTTCAGCCCTTTACTTCACGTTTGGAGATATACATGAG 1047
Db 482 GlyThrAlaThrAla-----AspAspIleLeuAsn 491
QY 1048 TTTCTGTGATGTTCTTAATAACTCGGAGCGGGTTTCGGATCTACATTTCAAGAGAGA 1107
Db 492 GlyValIleAsnAspAlaLysAspLysArgGlnAlaIleGluThrIleLeuGlnThrArg 511
QY 1108 GCATTAGAGCAAGAAACCCATCTCTGTGATCTTCTTCAATGGTGGATGATCGATAT 1167
Db 512 IleAsnLysAspLysAlaLysIleIleAlaAspValIleAlaArgValGlnLysAspLys 531
QY 1168 AGTCATTCGTTAGATGATTCATACGTTTATATACGGTTCCATGCTGCAACCGAGTTA 1227
Db 532 SerAspIleMetAspLeuIleHisSerAlaIleGluGly-----LysAlaAsnAspLeu 550
QY 1228 GATCCACAGTTACACACCCGGTTTCCCTCCAAACCGTTCTCTTATATACAAAGACCTG 1287
Db 551 AspIleGluLysArgAlaLysGlnAlaLysLysAspLeuGluTyIleLeuAspProIle 570
QY 1288 AGAGAGAGAACTGCAAGAGATATCTCTATGGATCTGTATTGGAGAGA---GGCAAA 1344
Db 571 LysAsnArg-----ProSerLeuLeuAspArgIleAsnLys 582
QY 1345 GACAGAGACTCAAGAAACCTGTATGTTCCACAGATTCCTCTTCAGCAGCTCAAAACGA 1404
Db 583 GlyValGlyAspSerAsnSerIlePheAspArgProSerLeuLeuAspLysLeuHisSer 602
QY 1405 AAGAACCATCAGATTGGAGACCTCAACGAGGTTTGCCTGAGAAATCTGTTTCGGTT 1461
Db 603 ArgGlySerIleLeuAspLysLeuAspHisSerAlaProGluAsnGlyLeuSerLeu 621
```

## RESULT 7

```
US-08-061-376-5
; Sequence 5, Application US/08061376
; Patent No. 6175000
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; APPLICANT: Djabali, Malek
; APPLICANT: Salleri, Lucia
; APPLICANT: Parry, Pauline
; TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11Q23
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pratty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/061,376
; FILING DATE: 13-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9387
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)546-4737
; TELEFAX: (619)546-9392
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3969 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-061-376-5
```

```
Alignment Scores:
Pred. No.: 0.0102 Length: 3969
Score: 112.50 Matches: 130
Percent Similarity: 31.50% Conservative: 76
Best Local Similarity: 19.88% Mismatches: 215
Query Match: 3.43% Indels: 235
DB: 3 Gaps: 26
```

US-09-423-575-1 (1-1886) x US-08-061-376-5 (1-3969)

```
QY 39 TCTGAATCTTTAGTGAGGAGATGATGAATGATGATTTCTTCATGAAATATTGTA 98
Db 2385 ThrGlnSerAlaAsnSerSerProAspGluAspThrGluVal----- 2398
QY 99 AGAAAAAAGCAATAGAGAGCTGCGGAATGAAAGTACACTGTTCTTTCACGGAGAAAGAA 158
Db 2399 -----LysThrLeuLysLeuSerGlyMetSer----- 2407
QY 159 GATAAATAGCATATCTCTCTTCAGTTTAAACACACATTTGGAAATTTTGATGTA 218
Db 2408 -----AsnArgSerSerIleIleAsnGluHisMetGly----- 2418
QY 219 AAAATCTCTTTGGAACGTTGTGTTCTGAAATCTTCCAAAGGTTCTATCAGAGAAG 278
Db 2419 -----SerSerSerArgAspArgArgGlnLys 2427
QY 279 AAGGATAAAGTTTCATAGAAACCCATGGACAAACAAACAAACAAACACACATTTAGTTC 338
Db 2428 GlyLysLysSerCys-LysGluThrPheLysGluLysHisSerSerLysSerPheLeuG 2447
QY 339 TCTGGATAATGTCATGACT-----AACCAAAATCTCTTCTCATGATTTAT 386
Db 2447 uProGlyGlnValThrThrGlyGluGluGlnLysProGluPheMetAspGluVa 2467
QY 387 ACCTTCAAGAGAGATTCAACTTCATTCACAAATGCTTCCATGGAATACCATCAATC 446
Db 2467 lLeuThrProGlu-----TyrMetGlyGlnArgProCysAsnValSerSe 2483
QY 447 AGAT-----CCTCTACA 458
Db 2483 rAspLysIleGlyAspLysGlyLeuSerMetProGlyValProLysAlaProPheMetG 2503
QY 459 AATCGGTGGC----- 468
Db 2503 nValGluGlySerAlaLysGluLeuGlnAlaProArgLysArgThrValLysValThrLe 2523
QY 469 -----TTTGATATTTCAT---TCTATGCTGCTACCAATATCTATCTCTCTCC 518
Db 2523 uThrProLeuLysMetGluAsnGluSerGlnSerLysAsnAlaLeuLysGluSerSerPr 2543
QY 519 ACGGTCT---ATCGATGTTCAAGATAACCGCAATGTTGAGTTTCATGGCTCTCTCTCTCA 575
Db 2543 cAlaSerProLeuGlnIleGluSerThrSerProThrGluProIleSerAlaSerGluAs 2563
```

QY	576	TCCT-----CCTCCACTTTCATCCCTTTGGATCAATTTAAGACACATATGATGATTCCTC	626
DB	2563	nProGlyaspGlyProvalAlaGlnProSerProAsnAsnThrSerCysGlnAspSerG1	2583
QY	627	AAACAACATGTGG-----	639
DB	2583	nSerAsnAsnTyrglnAsnLeuProvalGlnAspArgAsnLeuMetLeuProAspGlyPr	2603
QY	640	-----GGTTTGTGA-----GCAATAGTGA	659
DB	2603	oLysProGlnGluaspGlySerPheLysArgGlyProArgArgSerAlaArgAlaAr	2623
QY	660	GTTTCAGGCATTTTCAGGTAGTTGGTCCAAGTGAACCAATG-----ANGTCTACATT	713
DB	2623	gSerAsnMetPhePheGlyLeuThr-----ProLeuTyrglyValArgSerTy	2639
QY	714	CGGTGAAGAAGATTCCCGTTTCTTAATTTCGAAT-----AAAAGAAA	755
DB	2639	rGlyGluGluaspIleProPheTyrsSerSerThrGlyLysLysArgGlyLysArgSe	2659
QY	756	CAATGAGCTTTTCATGTAGTCTTGATCAGATGTTTCTTGATCAATGCTCGAG-----	807
DB	2659	rAlaGluGlyGlnValaspGlyAlaAspAspLeuSerThrSerAspGluaspLeuTy	2679
QY	808	-----ATAAGTCTTTGTGCAGCTACAAGATTAGCTCA-----	840
DB	2679	rTyTyAsnPheThrArgThrValIleSerSerGlyGlyGluGluArgLeuAlaSerHi	2699
QY	841	-----GAGCAAGCTTCTTCGACGAGCAAGACATTTCTTAATACGGTGTGTAC	887
DB	2699	sAsnLeuPheArgGluGluGlnCysaspLeuProLysIleSer-----	2714
QY	888	TCAAGGTTTCTCTCAACTATATTGGCTCAAAATACCTTCACCTCTGTCAAGAAATACT	947
DB	2714	-----	2714
QY	948	ATCTCATTTGCGCCATCTCGCTCGATTATTCATCTCGAGAACCGAGTCAGGAGTCG	1007
DB	2715	-----GlnLeuaspGlyValaspGlyThrGluSeraspThrSe	2728
QY	1008	TAGTTCAGCCTTTACTTCACGTTTTGAGATATAACTGAGTTTCTTCATGGTGATCTAA	1067
DB	2728	rValThrAlaThrThrArgLysSerSerGlnIleProLys---ArgAsnGlyLysGluAs	2747
QY	1068	TAATCTCGAGGC-----GGGTTTCGGATCTACATCTTC	1099
DB	2747	nGlyThrGluasnLeuLysIleaspArgProGluaspAlaGlyGluLysGluHisValTh	2767
QY	1100	AAAGGAGAGATTAAGACGAAGAAACCGATCTCTTGGATCTTCTTCAATGGTGGATG	1159
DB	2767	rLysSerSerValGlyHisLysasnGluProLys-MetAspAsnCysHisSerValSerA	2787
QY	1160	ATCGATATAGTTCGCTAGATGAGATTTCATACGGTTATATCAGCCTTC-----	1209
DB	2787	rgValLysThrGlnGlyGlnaspSerLeuGluAlaGlnLeuSerSerLeuGluSerSera	2807
QY	1210	-----CATGCTGCACCGAGTTAGATCCACAGTTTACACACCGGTTTGGCCCTCCAA-	1260
DB	2807	rgArgValHisThrSerThrProSerAspLysasnLeuLeuLeuLeuAspThrTyAsnThrGluL	2827
QY	1260	-----	1260
DB	2827	euleLysSeraspSeraspAsnAsnSeraspCysGlyAsnIleLeuProSerA	2847
QY	1261	--ACCGTTTCTCTTTATACAAGAACCTG-----AGAG	1291
DB	2847	spIleMetaspPheValLeuLysasnThrProSerMetGlnAlaLeuLysGluSerProG	2867
QY	1292	AGAGAATCTCGAAGAAGATAATCTCTATGGATCTGTATTGGAGAGAGGCAAGACAAGA	1351
DB	2867	luSerSerSerGluLeuLeuasnLeuGlyGluGlyLeu-----GlyLeuaspSera	2885

QY	1352	CTCAGAGAACTCTATGTTCCACGACGATTCCTTCTAGAGCTGAA-ACGAAAGAAC	1411
Db	2885	snArgGluLysAspMetGlyLeuPheGluValPheSerGlnGluLeuProThrGluP	2905
QY	1411	CATCAGATTTGGAGACCTCAACGAGGTTTGCCTGAGAAATCTCTTTCCGTT	1460
Db	2905	roValaspSerSerValSerSerIleSerAlaGluGluGlnPheGluLeuProLeuG	2925
QY	1461	-----TCTACGGAATTGGAGTT-----CCAAACTTCCTTACCCTTAC	1500
Db	2925	luLeuProSerAspLeuSerValLeuThrThrArgSerProThrValProSerGlnAsnP	2945
QY	1501	CCGAAAGATTCCGAGAGAACATCTTCTAGCTATACGAAG	1538
Db	2945	roSerArgLeuAlaValIleSerAspSerGlyGluLys	2957
RESULT 8			
US-09-449-285A-2			
; Sequence 2, Application US/09449285A			
; Patent No. 6313280			
; GENERAL INFORMATION:			
; APPLICANT: Vlaams Interuniversitair Instituut voor Biotechnologie			
; TITLE OF INVENTION: SMAD-INTERACTING POLYPEPTIDES AND THEIR USE			
; FILE REFERENCES: 2676-4232US			
; CURRENT APPLICATION NUMBER: US/09/449,285A			
; CURRENT FILING DATE: 1999-11-24			
; PRIOR APPLICATION NUMBER: PCT/EP98/03193			
; PRIOR FILING DATE: 1998-05-28			
; PRIOR APPLICATION NUMBER: 97201645.5			
; PRIOR FILING DATE: 1997-06-02			
; NUMBER OF SEQ ID NOS: 27			
; SOFTWARE: PatentIn version 3.0			
; SEQ ID NO 2			
; LENGTH: 944			
; TYPE: PRT			
; ORGANISM: Mus musculus			
US-09-449-285A-2			
Alignment Scores:			
Pred. No.:	0.00841	Length:	944
Score:	110.00	Matches:	106
Percent Similarity:	32.19%	Conservative:	73
Best Local Similarity:	19.06%	Mismatches:	178
Query Match:	3.35%	Indels:	199
DB:	4	Gaps:	28
US-09-423-575-1 (1-1886) x US-09-449-285A-2 (1-944)			
QY	195	ACACATTTGGAAATTTGATGTAAATTCCTTTGGAACTGTGTGTCTGAAATCT	254
Db	49	SerHisSerGlySerTyrSerSerHisIle-SerSerLysLysCysIle	64
QY	255	TCCAAAGGTTCTATC-----AGAAAGAAGAGATAAAAGTTTCATAGAAA	299
Db	65	-----GlyLeulleSerValasnGlyArgMetargasnIleLysThrGlySerSe	82
QY	300	CCCA-----ATGGACACACAAACAAACAACTTTAGTCTCTGGGATAATGTCAT	353
Db	82	rProasnSerValSerSerProThrAsnSerAlaIleThrGlnLeuAlaGlnLysLe	102
QY	354	GACTAACCAAAATCTCTTCTCATGGATTTTATACCTTCAAGAGAAGATTCAACTTCATT	413
Db	102	uGluasnGlyLysProLeu-----SerMe	110
QY	414	CTCAACAATTCCTCCATGGAATACCATCAGATCCTCTACAAATGGGTGGCTTGA	473
Db	110	tSerGluGlnThrGlyLeuLeuLysIleLysThrGluProLeuaspPheAsnAspTyrLy	130
QY	474	TATTTTCATTCCTGCTGACTAACAAATACTTATCATCTTCTCCACGGTCTATCGATGT	533
Db	130	valLeu-----MetAlaThrHisGlyPheSerGlySerPro	143
QY	534	TCAAGATAACCGCAATGTTGAGTTTCATG-----GTCCTCTCTCTCT	573

Db 144 -----PheMetAsnGlyGlyLeuGlyAlaThrSerProLeuG1 156  
QY 574 ----CATCTCTCCACTTCATCTTGGATCAITTAAGACACTATGATGATCTCAA 629  
Db 156 yValHisProSerAlaGlnSerProMetGlnHisLeuGly  
QY 630 CAACATGTGGGGTTTTGAAGCAAAATAGTGAGTTTCAGGCATTTTCAGGTGTAGTTGGTCC 689  
Db 170 -----ValGlyMe 172  
QY 690 AAGTGAACCAATGATGCTACATTCGGTGAAGAAGATTTCGGTTCTTAATTTTGAATAA 749  
Db 172 tGluAlaProLeuLeuGly-----PheProThrMetAsnSerAsnLe 186  
QY 750 PAGAACAATAGCTTTCATTCAGTCTTCATCAGTATGTTCTGATGAATGCTCGGAGAT 809  
Db 186 userGluValGlnLysValLeuGlnIleValAspAsnThr----- 199  
QY 810 AAGTCTTTGTGAGCTACAAGATTAGCTCTAGAGCAAGCTTCTTCAGACACAAGACAT 869  
Db 200 -----ValSerArgGlnLysMetAspCysLysThrGluAspI1 212  
QY 870 TTCT-----ATAACGTTGTTTACTCAAGTTTCTCTCACTTAT 908  
Db 212 eserLysLeuLysGlyTyrHisMetLysAspProCysSerGlnProGluGluGlnGlyVa 232  
QY 909 ATTTGGCTCAAATACTCTCACTGTTCAGAAATATCTATCTCATTTTCGCGCATACTC 968  
Db 232 lThrSerProAsnIleProProValGlyLeuProValSerHis----- 247  
QY 969 GCTCATATTTCATCTCAGGACCGAGTCAGGACTGCTAGTTCAGCTTACTTCTCAG 1028  
Db 248 -----AsnGlyAlaThrLysSerIleLeuAspTyrTh 258  
QY 1029 TTTTGAGATATAACTGAG-----TTTCTGTATGATGATTTCTAATAA 1070  
Db 258 rleuGluLysValAsnGluAlaLysAlaCysLeuGlnSerLeuThrThrAspSer----- 276  
QY 1071 CTCGAGCGGGTTTCGGATCTACATTTCAAAGGAGAGCATTAGAAGCAAGAAACCCA 1130  
Db 277 -----ArgArgGlnIleSerAsnIleLysLysGly 287  
QY 1131 TCTCTTGATCTTCTCAAAATGGTG--GATCATGATATATGTCATTCGGTAGATGAGAT 1187  
Db 287 sleuArgThrIleLeuAspLeuValThrAspAspLys-----MetIleGluAs 303  
QY 1188 TCATACGGTTATATCAGCGTTCCATGCT-----GCAACCGAGTTAGATCCA----- 1233  
Db 303 nHisSerIleSerThrProPheSerCysGlnPheCysLysGluSerPheProGlyProI1 323  
QY 1234 -----CAGTTACACACCGGTTT-----GCCCTCCAAACCGTTTC 1268  
Db 323 eProLeuHisGlnHisGluArgTyrLeuCysLysMetAsnGluGluIleLysAlaVal 343  
QY 1269 CTTCTTATACAGAACCTGAGAGAGAGA-----ATCTGCAAGAAGATAAT 1313  
Db 343 uGlnProHisGluAsnIleValProAsnLysAlaGlyValPheValAspAsnLysAlaLe 363  
QY 1314 CTCTATGGGATCTGATTGGAGAGGCGMAAGCAAGACTCAAGAAACCTCTATGTTCCA 1373  
Db 363 uLeuLeuSerSerValLeuSerGlu---LysGlyLeuThrSerProIleAsnProTyrLy 382  
QY 1374 CCAGGATTCCTTCTCAGCAGCTGAAACGAAAGAACCATCAGATTTGGAGACCTCAACG 1433  
Db 382 sAspHis----- 384  
QY 1434 AGGTTTGCCTGAGAAATCTGTTTCGGTTCTACGGAATTGGATGTTCCAAAACCTTCCTCA 1493  
Db 385 -----MetSerValLeuLysAlaTyrTyrAlaMetAsnMet----- 396  
QY 1494 CCTTACCCGAAAGATTCGGAGAAACAATCTTCTAGCTATACAGTGCCTTGACAGAG 1553

Db 397 ---GluProAsnSerAspGluLeuLeuLysIleSerIleAlaValGlyLeuProGlnG1 415  
QY 1554 TCAGGTATCAAACTGCTTTTATAAATCGGGGTTAGGCTATGGAAGCCGATGATAGAAGA 1613  
Db 415 uPheValLysGluThrPheGluGlnArgLysVal----- 426  
QY 1614 GATGTATCGGAAATCAACAAGAGG-----AAGCTCAATAACAGTCACATTC 1661  
Db 427 ---TyrGlnTyrSerAsnSerArgSerProSerLeuGluArgThrSerLysProLeuAl 445  
QY 1662 ACCAAC--GGACCAACTCTTCGA-----ATGCCAAATCT 1695  
Db 445 aProAsnSerAsnProThrThrLysAspSerLeuLeuProArgSer 460  
RESULT 9  
US-09-457-040B-27  
; Sequence 27, Application US/09457040B  
; Patent No. 6387641  
; GENERAL INFORMATION:  
; APPLICANT: Vertex Pharmaceuticals Incorporated  
; APPLICANT: Bellon, Steve  
; TITLE OF INVENTION: Crystallized P38 Complexes  
; FILE REFERENCE: VPI/98-14  
; CURRENT APPLICATION NUMBER: US/09/457,040B  
; CURRENT FILING DATE: 1999-12-08  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 27  
; LENGTH: 1584  
; TYPE: PRT  
; ORGANISM: DICTDI - Dictyostellium Discoideum  
US-09-457-040B-27  
Alignment Scores:  
Pred. No.: 0.0128 Length: 1584  
Score: 109.50 Matches: 105  
Percent Similarity: 29.98% Conservative: 68  
Best Local Similarity: 18.20% Mismatches: 187  
Query Match: 3.33% Indels: 217  
DB: 4 Gaps: 27  
US-09-423-575-1 (1-1886) x US-09-457-040B-27 (1-1584)  
QY 292 CATAGAAACCAATGGACACAAACAACAACACTTTTAGTTCTCTGGATAATGTC 351  
Db 494 AsnAsnAsnSerAsnSerAsnAsnAsnAsnAsnAsnAsnAsnAsnSerAsnSer 513  
QY 352 ATGACTAACCAAAATCCTCTTCTCATGGATTTTATACCTCAAGAGAGATTCACTCA 411  
Db 514 AsnSerAsnAsnAsnIleAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnIle 533  
QY 412 TTCTCAACATGCTTCCA-----TGGATACCATCAGATCCTCTACAATGGGT 465  
Db 534 TyrLeuThrLysProSerIleGlySerThrAspGluSerSerThrGlySerLeuGly 553  
QY 466 GGCTTTTGATATTTTCAATTCATGCTGCTAACAATACTATCATCTCTCCACGGTCT 525  
Db 554 Gly-----AsnAsnSerSerGlyAsnAsnSerSerSerGly----- 566  
QY 526 ATCGATGTTCAAGATACCCCAATGTTGAGTTTCAGTTCATCGCTCCTCTCTCAT- 576  
Db 567 ---SerIleGlyAsnSerSerIleLysGlnArgSerProHisSerIleAsn 585  
QY 577 ---CCTCCTCCATCTTCATCCTTTGGATCATTTAAGACACTATGATGATTCCTCAAAC 633  
Db 586 GlyProLeuMetLeuProPro-----SerSerThrAsnAsn 597  
QY 634 ATGTGGGGTTTGAAGCAAAATAGTGTTCAGGCATTT---TCAGGTGTAGTTGTCCA 690  
Db 598 -----AsnAsnAsnIleTyrSerSerTyrAsnSerThrThrAlaGlySer 612  
QY 691 AGTGAACCAATGATGTTCTACATTCGGTGAAGAGATTTCCTGTTTCTTAATTTCCGAATAA 750

613	SerThrThrLeuLeuProThrLeuAsnHisProIlePheGlyAsnThrThrSerAsn---	631
751	AGAAACAATGAGCTTTTCATTGAGTCTTGCATCAGAT-----GTTTCTGATGAATGC	801
632	AsnAsnSerSerSerThrLeuSerValGlyGlyAsnAsnAsnLeuGlyArgHisCys	651
802	TCGAGAGATAAGTCTTTGTGCGAGCTACAAGA-----	831
652	GlnSerLeuProIleThrAlaSerThrAsnHisThrLeuSerSerSerLeuGlyValSer	671
832	-----TTAGCCTCAGACCAAGCTCTTCTGCAGCAGCAACACATTTCTAAATAAGCTGTT	885
672	PheserSerProSerSerSerProLysThrSerProArgLysIleValAsnSerSerGlu	691
886	ACTCAAGGTTTC-----	897
692	AspLeuGlyPheValGlnThrPheGlnAspGlnProProSerAlaTrpArg	711
897	-----	897
712	ArgCysGlyLysSerIleLysThrLysAspAspIleThrLeuThrIleLysLysLys	731
898	-----TCTCAACTTATATTT	912
732	ThrSerValAlaMetAlaAspArgProPheSerSerAsnSerSerThrIleCysTrp	751
913	GGCTCAAAATACCTT-----CAC	930
752	PheGluValTyrLeuGluGlyHisAspLysGlySerIleThrValGlyLeuSerHis	771
931	TCGTGTTCAAGAAATACATATCTCATTTTCGCC-----GCATACTCGCTCGATTATTCATCT	984
772	SerThrTyrProPheIleLysHisIleGlyArgGluProLysSerTyrGlyPheSerSer	791
985	CGAGGAACCGAGTCAGAGCTGCTAGTTCAGCCTTACTTCACGTTTGTGAGATATTAAC	1044
792	GluGlyGluLysTyrGlyGlySer-----GluIleGly	802
1045	GAG-----TTCTTTGATGTGATTCTTAATAACTCGGAGGCGGGTTTC	1086
803	GluProTyrGlyProPhePhePheAspGlyAspSerIleAlaSerCysValIle	822
1087	GGA-----TCTACATTTCAAAGGAGAGCATTAGAAGCAAGAAACCCATCTCTTG	1137
823	GlyCysGlyIleAsnThrSerThrArgAspIlePhePheThrLysAsnGlyHisTyrLeu	842
1138	GATCTTCTTCAAATGGTGGATGTCGATAGTATGATTCGCTAGATGAGATTCATCGGTT	1197
843	GlyVal-----	844
1198	ATATCAGCGTTCCATGTCGCAACCGAGTTAGATCCACAGTTACACACCCCGTTTGCCCTC	1257
845	-----AlaPheSerArgValThrSer--AspPro-----	853
1258	CAAAACGGTTTCCCTCTTATACAGAACTCAGAGAGAGA-----ATCTGC	1302
854	-----LeuTyrProSerIleSerPheArgGlyValValGlyGlyLeuCys	868
1303	AAGAAGATAATCTCTATGGGATCTGTTATTGGAGAGAGCAACAGACTCAAGAAACC	1362
869	ValAlaThrPheProGlyGly-----	875
1363	TCTATGTTCCACGAGCATTGCCCTCTTTCACAGCTGAACAGAAACCATCAGATTGG	1422
876	-----HisPheArgPheAsnIleGluAspLeuProGlyIleSerProSerValTrp	892
1423	AGACCTCAACGAGGTTTGCTCGAATAATCTGTTTCGGTTCTACGGAAATTCG-----	1473
893	ThrGluAlaLeuGly---ProAspArgGlnGlySerGlyPheLysAsnTrpAlaProAsn	911
1474	-----ATGTTCCAAACTTCCTTCAC-----CCTTACCCGGAAGAT-----	1509
912	AspValAlaIleTrpLeuGluSerPheAsnTyrGlyGlnTyrArgLysAsnPheArgAsp	931

```

1510  QY  -----TCGAGAGAAACATCTTCTAGCTATACGAAAGTGGCTTGACAAGACTCAGGTA 1560
      |||  :|||:|||||
932   Db  AsnAsnIleSerGlyArgHisLeu-----GluglyIleThrHisalaMetLeu 947
      |||  :|||:|||||
1561  QY  TCAAACTGGTTT-----ATAAAT 1578
      |||  :|||:|||||
948   Db  LysAsnAspLeuGlyIleGluProTyrGlyHisArgGluAspIleIleAsnArgLeuAsn 967
      |||  :|||:|||||
1579  QY  GCDCGGGTAGGCTATGGAAGCGCATGATAGAGAGATATATCGGAAATG 1629
      :|||:|||||
968   Db  ArgMetIleGlnIleTrpAsnAspLysSerProAspSertyrProlysile 984
      :|||:|||||

RESULT 10
US-09-107-532A-3678
; Sequence 3678, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: LYVIN A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Atinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3678:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3) LOCATION 1...522
; SEQUENCE DESCRIPTION: SEQ ID NO: 3678:
US-09-107-532A-3678

Alignment Scores:
Pred. No.: 0.00993 Length: 522
Score: 108.00 Matches: 86
Percent Similarity: 38.8% Conservative: 81
Best Local Similarity: 20.0% Mismatches: 149
Query Match: 3.29% Indels: 114
DB: 4 Gaps: 22

US-09-423-575-1 (1-1886) x US-09-107-532A-3678 (1-522)

```

QY	807	GATAAGTCTTTGTGCAGCTACAAGATTAGCCTCAGAGCAAGCTTCTTTGCAGCAGCAAGA	866
		::::	::::: ::::
Db	90	AspProAlaIatyrCysTyrTyr-ArgLeuPheGlyAArgTyrCys-	106
QY	867	CATTTCCTAATAACGTTGTTTACTCAAGGTTTCTCT-	900
		::::	-----
Db	106	rValSerTyrGlnIleuThrSerLeuPheSerCysGlnThrAsnSerIleIleSerLe	126
QY	901	-----CAACTTATATTGGCTCAAAATAACCTTCACTCTGTTTCAAGAAATATCTATCTCA	953
		::: ::::	::::: ::::
Db	126	uSerGlnGlnThrAsnTyr-SerAlaSerTyrLeuTyrThrLysMetLysLysIleAsnAl	146
QY	954	TTTCGCCGCA-----TACTCGCTCGATTATTCATCTCGAGGACACCGAGTCA-----	999
		::::	::::: ::::
Db	146	aPheLeuAlaLeuTyrGlyLeuSerIleSerPheSerAsnAsnGlyAArgLysSerPheSe	166
QY	1000	-----GGAGCTGTAGTTTTCAGCTTTTACTTCTACTCGTTTTCAGAA	1037
		::::	::::: ::::
Db	166	rGlyLysGluValGlnIleGlnTyrAlaAlaLeuAspIleTyrTrpAsnIlePheSerSe	186
QY	1038	T-----ATAACTAGTCTTCTGATGGTATCTCTAATACTCGAGCGGGTTCGGAATCTAC	1094
		::::	::::: ::::
Db	186	rThrAlaThrGluPheTyrAspGluAspProGlnValValAlaPheMetMetAsnThrPh	206
QY	1095	ATTTCAAAGAGAGAGCATTAGAAGCAAGAAACCCATCTCTTGGAT-----CT	1142
		::::	-----
Db	206	ePheLysGluValLeuAspArgLeuAsnSerGlyMetIleAspLysLeuTyrPheLe	226
QY	1143	TCTTCAAATGGTGGATCATCGATAT-----AGTCATTGGTATAGATCAGATTCATACGGTTAT	1199
		::::	::::: ::::
Db	226	uLeuLysIleCysLysGluAsnPheProAlaThrSerValLysAspValGlnLysGluPh	246
QY	1200	ATCAGCGCTTC---CATGCTGCAACCGAGTTAGATCCACAGTTACACACCGGTTTGCCCT	1256
		::::	-----
Db	246	eThrGluTyrGluTyrIleAspPheIleAspPro-----	258
QY	1257	CCAAACGGTTCTCTTATACAAAGACCTGAGAGAGAGATCTGCNAGAGATATATCTC	1316
		::::	::::: ::::
Db	259	---SerValAspIleLeuArgProAsnLeu-----ProIleCysArgGluGlnArgIl	275
QY	1317	TATGGGATCTGTATTGGAGAGA-----GGCAAGACACAAGACTCAAGAAACCTCTAT	1367
		::::	::::: ::::
Db	275	eLeuValAsnIleLeuAlaAlaArgLeuSerIleThrLysIleGluSerGluGluThrSerLe	295
QY	1368	GTTCACCAACAGATTCCTCTTCTCAGCAGCTGAACAGNAAAGAAC-----CATCAGATTG	1421
		::::	::::: ::::
Db	295	u---GluGlnTyrLysLeu-----LeuLysLysAlaAsnValProHisValLeuTy	311
QY	1422	GAGACCTCAACGA-----GGTTTGCTCTCAGAAAA--	1449
		::::	-----
Db	311	rSerArgGluMetValGluSerPheCysThrValPheAspLeuPheIleProGluLysGl	331
QY	1450	-----TCTGTTTCGGTTCTACGGAATGGATGTTCACAAACTCTCTCTAC--	1494
		::::	::::: ::::
Db	331	uLysIleLeuTyrSerLeuAlaPheAlaArgAsnLysLeuTyrAsnGlyPheLeuAsnPh	351
QY	1495	-----CCTTACCCG-----	1503
		::::	-----
Db	351	eAsnGlnProAsnThrProLeuProThrPheMetLeuTyrGlnGluHisSerValLeuLy	371
QY	1504	-----AAGATTGGAGAAAAATCT	1523
		::::	::::: ::::
Db	371	sGluValGlnThrAlaIleGluGlnPheTyrMetAspPheArgGluGlnAsnGlnHisLe	391
QY	1524	TCTAGCTATACGAAGTGGCTTGACAAGAGTCAGGTATCAAACCTGGTTTATAATGCGCG	1583
		::::	::::: ::::
Db	391	uLeuProVal-----IleLeuGluLysGluAsnValGlnTrpMetValGluAspLe	408
QY	1584	GGTTAGGCTATGGAAGCCGATGATAGAAGAGATGATATGCGGAATGAACAGAGGAGACT	1643
		::::	::::: ::::
Db	408	uValHisLeu-----TyrAspArgTyrLysLysGln-ProA	420
		::::	::::: ::::

Qy	1644	CATATAAGTCACATTCAACCCAAACGGAACAACCTCTTGCRAATGCCAAAATCTGTATGAT	1705
Db	420	rgilevalilleGlyValAsnTyThrArgAspPheTyrlleSerFlysAspLeuileVall	440
Qy	1704	GAGCCAAGCAATGCATATAAATAAGACACAATTTGGTTTACCACAC---TTTGTCGATAATTA	1760
Db	440	yslleGluGlnIlePheSerAsnGluSerileLyslleGlnAsnAspPheMetGluAsnC	460
Qy	1761	GGCAATTGCTACTCTATGATGGCCCA---AAACCTAAACCATGTACGACTATCATTACGT	1817
Db	460	ysAnllleValilleSerAspCyseProleuLysGlnLeuProserHisileLysLysileT	480
Qy	1818	ATGTTATAATTGTATATACAACTCTCT	1843
Db	480	yrilleLeuLysGlylleThrPro	488
 RESULT 11 US-08-194-991E-5 ; Sequence 5, Application US/08194981E ; Patent No. 5886157 ; GENERAL INFORMATION: ; APPLICANT: GUENTERICH, F. Peter ; APPLICANT: GUO, Zuyu ; APPLICANT: SANDHU, Punam ; APPLICANT: GILLAM, Elizabeth M. J. ; TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF ; TITLE OF INVENTION: HUMAN ; TITLE OF INVENTION: CYTOCHROME P450 ; NUMBER OF SEQUENCES: 68 ; CORRESPONDENCE ADDRESS: ; ADDRESSEE: NEEDLE & ROSENBERG, P.C. ; STREET: Suite 1200, 127 Peachtree Street, NE ; City: Atlanta ; STATE: Georgia ; COUNTRY: USA ; ZIP: 30303-1811 ; COMPUTER READABLE FORM: ; MEDIUM TYPE: Floppy disk ; COMPUTER: IBM PC compatible ; OPERATING SYSTEM: PC-DOS/MS-DOS ; SOFTWARE: Patentin Release #1.0, Version #1.25 ; CURRENT APPLICATION DATA: ; APPLICATION NUMBER: US/08/194,981E ; FILING DATE: February 10, 1994 ; CLASSIFICATION: 435 ; ATTORNEY/AGENT INFORMATION: ; NAME: Elizabeth Selby ; REGISTRATION NUMBER: 38,298 ; REFERENCE/DOCKET NUMBER: 22000.0022 ; TELECOMMUNICATION INFORMATION: ; TELEPHONE: (404) 688-0770 ; TELEFAX: (404) 688-9880 ; INFORMATION FOR SEQ ID NO: 5: ; SEQUENCE CHARACTERISTICS: ; LENGTH: 512 amino acids ; TYPE: amino acid ; STRANDEDNESS: single ; TOPOLOGY: linear ; MOLECULE TYPE: peptide ; HYPOTHEetical: NO ; FRAGMENT TYPE: N-terminal US-08-194-991E-5			
 Alignment Scores:			
Pred. No.:	0.0126	Length:	512
Score:	107.00	Matches:	94
Percent Similarity:	34.38%	Conservative:	59
Best Local Similarity:	21.12%	Mismatches:	170
Query Match:	3.26%	Indels:	122
DB:	2	Gaps:	23
 US-09-423-575-1 (1-1886) x US-08-194-981E-5 (1-512)			

427 CCATGAATACCATCAGATCAGATCCTCTCAAAATGGGTGGTCTTGATATTTCATTTCT 486  
 Db |||||  
 43 ProTrp -----GlyTrpProLeuIleGlyHis 51  
 QY |||||  
 487 ATGCTGACT-----AACAAATACTATCATCTTCTCCACGGTCTCATGATGTTCAA 537  
 Db |||||  
 52 MetLeuThrLeuGlyLeuAsnProHisLeuAlaLeuSerArgMetSerGlnGlnTyrgly 71  
 QY |||||  
 538 GATAACCGCAATGTTGAGTTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 597  
 Db |||||  
 72 AspValLeuGlnIleArgIleGlyThrPro-----ValValValLeuSerGlyLeu 88  
 QY |||||  
 598 GATCATTTTAAGACACTAT-----GATGATTCTCTCAAAACAATGTTGGGT 642  
 Db |||||  
 89 AspThrIleArgGlnAlaLeuValArgGlnGlyAspAsp----- 101  
 QY |||||  
 643 TTTGAAGCAATAGTGAGTTCAGGCATTTTCAGGTGAGTGTGTCAGAGTGAACCAATG 702  
 Db |||||  
 102 PheLysGlyArgProAspLeuTyThrPheThrLeuIle-----SerAsnGlyGln 118  
 QY |||||  
 703 ATGCTCATCTGGTGAAGAAGATTTCCTGTTCTTAATTTTCAATAAAAGA----- 753  
 Db |||||  
 119 SerMetSerPheSerProAspSerGlyProValTrpAlaAlaArgArgLeuAlaGln 138  
 QY |||||  
 754 RACATGAGCTTTCATTCAGTCTTCATCAGATGTTTCTGATGATGCTCGAGATAGT 813  
 Db |||||  
 139 AsnGlyLeuLysSerPheSerIleAlaSerProAlaSerThrSer----- 155  
 QY |||||  
 814 CTTTGTGAGCTACAGATTAGCTTCAGAGCAAGCTTCTTCAGCAGCAAAAGACATTCT 873  
 Db |||||  
 156 ---CysTyLeuGluGluHisValSerLysGluAlaGluValLeuLeuSerThrLeuGln 174  
 QY |||||  
 874 AATAAGCT-----GTTACT 889  
 Db |||||  
 175 GluLeuMetAlaGlyProGlyHisPheAsnProTyrArgTyrValValValSerValThr 194  
 QY |||||  
 889 CAAGGTTCTCTCAACTATATTGGCTCAAAATACCTTCACTCTCTGTTCAAGAAATACTA 948  
 Db |||||  
 195 AsnValIleCysAlaIleCysPheGlyArgTyrAspHisAsnHisGlnGluLeuLeu 214  
 QY |||||  
 949 TCTATTTCCGCGATATCTGCTGATTTATTCATCTCGAGAACCGAGTCAAGAGTGCT 1008  
 Db |||||  
 215 Ser-----LeuValAsnLeuAsn---AsnAsnPheGlyGluValValGlySerGly 230  
 QY |||||  
 1009 AGTTGAGCTTTTACTTCACTGAGTTCAGAAATATACTGATGTTCTTGTGTTGATTCTAAT 1068  
 Db |||||  
 231 AsnProAla-----GluPheIleProIleLeuArgTyrLeuProAsnProSerLeu 247  
 QY |||||  
 1069 AACTCG-----GAGCGCGGTTTCGGATCTCATCTTCAAGAGGAGCATTAGAA 1116  
 Db |||||  
 248 AsnAlaPheLysAspLeuAsnGlnLysPheTyrSerPheMetGlnLysMetValLysGlu 267  
 QY |||||  
 1117 GCAAGAAACCC-----CATCTCTGATCTTCTTCAATGGTGGATGATCGA 1164  
 Db |||||  
 268 HisTyrLysThrPheGluLysGlyHisIleArgAspIle-----ThrAspSerLeu 284  
 QY |||||  
 1165 TATAGTCATTGC-----GTAGATGAG----- 1185  
 Db |||||  
 285 IleGluHisCysGlnGluLysGlnLeuAspGluAsnAlaAsnValGlnLeuSerAspGlu 304  
 QY |||||  
 1186 -----ATTATACGGTTATATCAGGTTTCATCTGCAACCGAGTAGATCCACAGTTA 1239  
 Db |||||  
 305 LysIleIleAsnIleValLeuAspLeuPheGlyAlaGlyPheAsp----- 319  
 QY |||||  
 1240 CACACCCGGTTGCGCTCCAAACCGTTCTCTTATACAGAACCTGAGAGAGAGATC 1299  
 Db |||||  
 320 ---ThrValThrAlaIleSerTrpSerLeuMetTyrLeuValMetAsnProArgVal 338  
 QY |||||  
 1300 TGCAAGAGATA---ATCTTATGGATCTCTATTGGAGAGCAAGCAAGACTCAA 1356  
 Db |||||  
 339 GlnArgLysIleGlnGluLeuAspThrValIleGlyArgSerArg----- 354  
 QY |||||  
 1357 GAAACCTCTATGTTCCACCAAGATTGCGCTTCTTTCAGCAGCTGAAACCAAGAACCATCAG 1416

355 -----ArgProArgLeuSerAspArgSerHis--- 363  
 Db |||||  
 1417 ATTTGGAGACCTCAACGAGGTTTGCTGAGAAATCTGTTTTCGTTCTACGGAATTGGATG 1476  
 QY |||||  
 364 -----LeuProTyrMetGluAlaPheIleLeuGluThrPheArg 376  
 Db |||||  
 1477 TTCAAAATCTTCCT-----CACCCCTTACCCGAAAGATTTCGGAGAAACAT 1521  
 QY |||||  
 377 HisSerSerPheValProPheThrIleProHisSerThrThrArgAspThrSerLeuLys 396  
 Db |||||  
 1522 CTCTAGCTATACGAAGTGGCTTGACAAAGATCAGGTATCAAACTGGTTTATAATGCG 1581  
 QY |||||  
 397 GlyPheTyIleProLysGlyArgCys---ValPheValAsnGlnTrpGlnIleAsnHis 415  
 Db |||||  
 1582 CGGGTTAGGCTATGG 1596  
 QY |||||  
 416 AspGlnLysLeuTrp 420  
 Db |||||  
 RESULT 12  
 US-08-480-662-2  
 ; Sequence 2, Application US/08480662  
 ; Patent No. 5759782  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Pastan, Ira  
 ; APPLICANT: Brinkmann, Ulrich  
 ; TITLE OF INVENTION: CELLULAR APOPTOSIS SUSCEPTIBILITY PROTEIN (CSP) AND AN  
 ; NUMBER OF SEQUENCES: 10  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Knobbe, Martens, Olson and Bear  
 ; STREET: 620 Newport Center Drive 16th Floor  
 ; CITY: Newport Beach  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 92660  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq Version 1.5  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/480,662  
 ; FILING DATE: 07-JUN-1995  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Israel, Ned A  
 ; REGISTRATION NUMBER: 29,655  
 ; REFERENCE/DOCKET NUMBER: NH112.001A  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 619-235-8550  
 ; TELEFAX: 619-235-0176  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 971 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; FRAGMENT TYPE: N-terminal  
 ; ORIGINAL SOURCE:  
 ; US-08-480-662-2  
 Alignment Scores:  
 Pred. No.: 0.0207 Length: 971  
 Score: 106.50 Matches: 117  
 Percent Similarity: 34.05% Conservative: 74  
 Best Local Similarity: 20.86% Mismatches: 193



Query Match:	3.24%	Indels:	177
DB:	1	Gaps:	26
US-09-423-575-1 (1-1886) x US-08-480-662-2 (1-971)			
QY	181	CTTCGTTTAAACACACATTTTGGAAATTTGATGTAATAATCTCTTTGGAACGTTGT	240
DB	8	LeuGlnThrLeuThrGluThrLeu	20
QY	241	GTTGCTGAAATCTTCCCAAAGGTTCTATCAGAAGAAGAAGATAAAGTTTTCATAGAAC	300
DB	21	ProAspProAlaIleArgAspProAlaGluLysPhe	32
QY	301	CCAATGGACAACAACAACAACACATTTTAGTCTCTGGATAATGTCATCACTAAC	360
DB	33	LeuGluSerValGluGlyAsn	39
QY	361	CAAAT-----CCTCTTCTCATGATTTTATACCTTCAAGAGAAGAT-----	402
DB	40	GlnAsnTyrProLeuLeuLeuLeuThrLeuLeuGluLysSerGlnAspAsnValIleLys	59
QY	403	-----TCAACTTCATCTCAACAATGCTTCCA-----TGAATACCATCATGATCA	447
DB	60	ValCysAlaSerValThrPheLysAsnTyrIleLysArgAsnTrpArgIleValGluAsp	79
QY	448	GATCCCTCAAAATGGTGGCTTTGAT-----ATTTTCAATCT	486
DB	80	GluProAsnLysIleCysGluAlaAspArgValAlaIleLysAlaAsnIleValHisLeu	99
QY	487	ATGCTGACTACAAATPACTTATCATCTTCCACGGTCTATCGATGTTTCAGATAACCGC	546
DB	100	MetLeu-----SerSerProGluGlnIleGlnLysGlnLeuSerAsp	113
QY	547	AATGTTGAGTTCATGCTCCTCTCATCTCTCTCCCTCCATCTCATCTTCTGGATCATTTA	606
DB	114	AlaIleSerIleIleGlyArgGluAspPhePro-----	124
QY	607	AGACATATGATGATTCCTCAACAACATGTGGGTTTGAAGCAAAATAGTCAGTTTCAG	666
DB	125	GlnLysTrpProAspLeuLeuThrGluMetValAsnArgPheGlnSerGlyAspPheHis	144
QY	667	GCATTTTCAGCTAGTGTGTCAGTGAACCAATGATGTCTACATTCGGTGAAGAGAT	726
DB	145	ValIleAsnGlyValLeuArgThrAlaHisSerLeuPheLysArgTyrArgHisGlu---	163
QY	727	TTCCCGTTCTAATTTTCGAATAAAGAAACAATGAGCTTTCATTCAGCTTTCGCATCAT	786
DB	164	-----PheLysSerAsnGluLeuThrGluIleLysLeuValLeu-----	177
QY	787	GTTTCTGATGATGTCGGAGATAAGTCTTTGTGAGTACAGATTAGCCTCAGAGCAA	846
DB	178	-----AspAlaPheAlaLeuProLeuThrAsnLeuPheLysAlaThrIleGluLeu	194
QY	847	GCTTCTGCGACGACAAACATTTCTAATAACGTTGTTACTCAAGGTTTCTCAACTT	906
DB	195	CysSerThrHisAlaAsnAspAlaSerAlaLeuArgIleLeu-----PheSerSerLeu	212
QY	907	ATATTTGGCTCAAAATACCTTCACTCTGTTCAAGAAATACTATCTATTCCTTCGCGCAT	966
DB	213	IleLeuIleSerLysLeuPheTyrSerLeu-----	222
QY	967	TCGCTCATATTTCATCTCGAGAACCGAGTCAGGCTGCTAGTTCAGCCTTTACTTCA	1026
DB	222	-----	222
QY	1027	CGTTTTCAGAAATATACTAGTGTCTTGTGATGATTTCTAATACTCGAGCGGGTTTC	1086
DB	223	AspPheGlnAspLeuProGluPheTrpGluGlyAsnMetGluThrTrpMetAsnAsnPhe	242
QY	1087	GGATCT-----ACATTTCAAAGGAGCATTAAGAACA-----AAGAAACCCAT	1131
DB	243	HisThrLeuLeuThrLeuAspAsnLysLeuLeuGlnThrAspGluGluAlaGly	262

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NH112.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-235-8550
; TELEFAX: 619-235-0176
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 971 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; US-08-918-190-2
;
; Alignment Scores:
; Pred. No.: 0.0207 Length: 971
; Score: 106.50 Matches: 117
; Percent Similarity: 34.05% Conservative: 74
; Best Local Similarity: 20.86% Mismatches: 193
; Query Match: 3.24% Indels: 177
; DB: 26
;
; US-09-423-575-1 (1-1886) x US-08-918-190-2 (1-971)
;
; QY 181 CTTTCAGTTTAAACACACATTTTGGAAATTTTGATGTAAATAATCTCTTTGGAAGTTGT 240
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;   8 LeuGlnThrLeuThrGluThrLeu-----LysLysThrLeuAsp----- 20
;
; QY 241 GTTGCTGAAATCTCCCAAGGTTCTATCAGAGAAGAGGATAAAGTTTCATAGAAAC 300
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;   21 -----ProAspProAlaIleArgProAlaGluLysPhe----- 32
;
; QY 301 CCAATGGACACAAACACAAACACACTTTTAGTCTCTGGATAAGTCTATGACTAAC 360
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;   33 -----LeuGluSerValGluGlyAsn 39
;
; QY 361 CAAAT-----CCTCTTCTCATGGATTATPACCTTCAAGAGAGAT----- 402
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;   40 GlnAsnTyrProLeuLeuLeuThrLeuLeuGluLysSerGlnAsnValIleLys 59
;
; QY 403 -----TCAACTTCATCTCAACAATGCTTCCA-----TGAATACCATCAGATCA 447
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;   60 ValCysAlaSerValThrPheLysAsnTyrIleLysArgAsnTrpArgIleValGluAsp 79
;
; QY 448 GATCCTCTCAAAATGGGTGGCTTGCAT-----ATTTTCAATTCT 486
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;   80 GluProAsnLysIleCysGluAlaAspArgValAlaIleLysAlaAsnIleValHisLeu 99
;
; QY 487 ATGCTGCAACAAATACTATCATCTCTCCACGGTCTATCGATGTTCAAGATAACCGC 546
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;   100 MetLeu-----SerSerProGluGlnIleGlnLysGlnLeuSerAsp 113
;
; QY 547 AATGTTGAGTTCATGGCTCCTCCTCCTCATCCTCCATCTCATCTTGGATCAATTA 606
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;   114 AlaIleSerIleIleGlyArgGluAspPhePro----- 124
;
; QY 607 AGACACTATGATGATCTCTCAAAACAACATGTGGGGTTTTGAAGCAAAATAGTGGTTTCAG 666
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;   125 GlnLysTyrProAspLeuLeuThrGluMetValAsnArgPheGlnSerGlyAspPheHis 144
;
; QY 667 GCATTTTCAGGTGTAGTTGGTCCCAAGTGAACCAATGATGTCTACATTCGGTGAAGAAGAT 726
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;   145 ValIleAsnGlyValIleuArgThrAlaHisSerLeuPheLysArgTyrArgHisGlu--- 163
;
; QY 727 TTCCCGTTTCTTAATTCGAATAAAGAAACAATGAGCTTTTCATTGAGCTTTTCATCAGAT 786
;   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Patent No. 6232086  
 GENERAL INFORMATION:  
 APPLICANT: Paetan, Ira  
 TITLE OF INVENTION: CELLULAR APOPTOSIS SUSCEPTIBILITY  
 PROTEIN (CSP) AND ANTISENSE CSP  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Krobbe, Martens, Olson and Bear  
 STREET: 620 Newport Center Drive 16th Floor  
 CITY: Newport Beach  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 92660

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq Version 1.5  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/234,232  
 FILING DATE: 20-Jan-1999  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/919,190  
 FILING DATE: <Unknown>  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Israelsen, Ned A  
 REGISTRATION NUMBER: 29,655  
 REFERENCE/DOCKET NUMBER: NIH112.001A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 619-235-8550  
 TELEFAX: 619-235-0176  
 TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 TYPE: amino acid  
 LENGTH: 971 amino acids  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FRAGMENT TYPE: N-terminal  
 ORIGINAL SOURCE:  
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-234-232-2

Alignment Scores:  
 Pred. No.: 0.0207 Length: 971  
 Score: 106.50 Matches: 117  
 Percent Similarity: 34.05% Conservative: 74  
 Best Local Similarity: 20.86% Mismatches: 193  
 Query Match: 3.24% Indels: 177  
 DB: 3 Gaps: 26

US-09-423-575-1 (1-1886) x US-09-234-232-2 (1-971)

QY	181	CTTCAGTTTAAACACACATTTTGGAAATTTGATGTAATAATCTCTTTGGAACGTTCT	240
Db	8	LeuGlnThrLeuThrGluTyrLeu	20
QY	241	GTTCGTGAAATCTCCAAAGGTTCTATCAGAGAGAGAGGATAAGTTTCATGAAC	300
Db	21	-----ProAspProAlaIleArgArgProAlaGluLysPhe	32
QY	301	CCAATGGACAACACACACACACATTTTAGTCTCTGATATGTCATGACTAAC	360
Db	33	-----LeuGluSerValGluGlyAsn	39
QY	361	CAAAAT-----CCTCTTCTCATGGATTTTATACCTTCAAGAGAAGAT	402
Db	40	GlnAsnTyrProLeuLeuLeuThrLeuLeuGluLysSerGlnAspAsnValIleLys	59

QY	403	-----TCAACTTCATCTCAACAATGCTTCCA-----TGAATACCATCATCATCA	447
Db	60	ValCysAlaSerValThrPheLysAsnTyrIleLysArgAsnTrpArgIleValGluAsp	79
QY	448	GATCCCTCAAAATGGTGGCTTTGAT-----ATTTTCAATTCT	486
Db	80	GluProAsnLysIleCysGluAlaAspArgValAlaIleLysAlaSerIleValHisLeu	99
QY	487	ATGTGACTAACAATACTTATCATCTTCTCCACGGTCTCTATCGATGTTCAAGTAACCGC	546
Db	100	MetLeu-----SerSerProGluGlnIleGlnLysGlnLeuSerAsp	113
QY	547	ATGTTGAGTTTCATGGCTCTCTCTCCATCTCTCCATCTCTCCATCTCTCTTGGATCATTTA	606
Db	114	AlaIleSerIleIleGlyArgGluAspPhePro-----	124
QY	607	AGACACTATGATGATCTCTCAACAACATGGGGTTTGAAGCAAAATAGTGTTCAG	666
Db	125	GlnLysTrpProAspLeuLeuThrGluMetValAsnArgPheGlnSerGlyAspPheHis	144
QY	667	GCATTTTCAGGTAGTGGTCCAAAGTGAACCAATGATGCTCATCTCGGTGAAGAGAT	726
Db	145	ValIleAsnGlyValLeuArgThrAlaHisSerLeuPheLysArgTyrArgHisGlu---	163
QY	727	TTCCCGTTTCTAATTCGAATAAAAGAAACAATGAGCTTTTCATTGAGTTCATGATCAGAT	786
Db	164	-----PheLysSerAsnGluLeuTrpThrGluIleLysLeuValLeu-----	177
QY	787	GTTTCTCATGATGCTCGGAGATAAGTCTTTGTGCTAGCTACAAGATTAGCCTCAGACAA	846
Db	178	-----AspAlaPheAlaLeuProLeuThrAsnLeuPheLysAlaThrIleGluLeu	194
QY	847	GCTTCTTCGACGACCAAGACATTTCTAATAACGTTCTTACTCAAGTTCCTCTCAACTT	906
Db	195	CysSerThrHisAlaAsnAspAlaSerAlaLeuArgIleLeu-----PheSerSerLeu	212
QY	907	ATATTGGCTCAAAATACCTTCACTCTCTCTCAAGAAATACCTATCTCTTTCGCGCATC	966
Db	213	IleLeuIleSerLysLeuPheTyrSerLeu-----	222
QY	967	TCGCTCGATATTATCTCTCGAGAACCGAGTCAGAGCTGCTAGTTCAGCCTTACTTCA	1026
Db	222	-----	222
QY	1027	CGTTTGGAGATATACTAGTGTTCGATGTTCTTATGATGTTCTTAATACTCGGAGCGGTTTC	1086
Db	223	AsnPheGlnAspLeuProGluPheTrpGluGlyAsnMetGluThrTrpMetAsnAsnPhe	242
QY	1087	GGATCT-----ACATTTCAAAGGAGCATTAGAACCA-----AGAAAACCCAT	1131
Db	243	HisThrLeuLeuThrLeuAspAsnLysLeuLeuGlnThrAspAspGluGluAlaGly	262
QY	1132	CTCTTGGATCTCTT-----CAATGGTG	1155
Db	263	LeuLeuGluLeuLeuLysSerGlnIleCysAspAsnAlaAlaLeuTyrAlaGlnLysTyr	282
QY	1156	GATGATCATATAGTATGCTAGATGAGATTATACGGTTATATACGGTTCATGCT	1215
Db	283	AspGluGluPheGlnArgTyrLeuProArgPheValThrAlaIleTrpAsnLeuVal	302
QY	1216	GCAACC-----GAGTAGATCCACAGTTACACACCGGTTTGGCCTCCAAACCGTTTCC	1269
Db	303	ThrThrGlyGlnGluValLysTyrAspLeuLeuValSerAsnAlaIleGlnPheLeuAla	322
QY	1270	TTCTTA-----TACAAGAACCTGAGAGAG-----AGA	1296
Db	323	SerValCysGluArgProHisTyrLysAsnLeuPheGluAspGlnAsnThrLeuThrSer	342
QY	1297	ATCTCCACAGAGATAATCTCTATGGGATCTCTATTTGNGA-----GAGGCAAA	1344
Db	343	IleCysGluLysValIle-ValProAsnMetGluPheArgAlaAlaAspGluGluAla-P	362



Db 223 AsnPheGlnAspLeuProGluPheTrpGluGlyAsnMetGluThrTrpMetAsnAsnPhe 242  
QY 1087 GGATCT-----ACATTTCAAGGAGAGCATAGAGCA-----AAGAAAACCCAT 1131  
Db 243 HisThrLeuLeuThrLeuAspAsnLysLeuGlnThrAspAspGluGluGluAlaGly 262  
QY 1132 CTCTTGGATCTCTT-----CAAAATGGTG 1155  
Db 263 LeuLeuGluLeuLeuLysSerGlnIleCysAspAsnAlaAlaLeuTyAlaGlnLysTyr 282  
QY 1156 GATGATCGATATPAGTCATTGCGTATGATGATTCATCGTTTATATCAGCGTTCCATGCT 1215  
Db 283 AspGluGluPheGlnArgTyrLeuProArgPheValThrAlaIleTrpAsnLeuVal 302  
QY 1216 GCATCC-----GAGTTAGATCCACAGTTACACACCGGTTTCCCTCCAAACCGTTTCC 1269  
Db 303 ThrThrGlyGlnGluValLysTyrAspLeuLeuValSerAsnAlaIleGlnPheLeuAla 322  
QY 1270 TTCTTA-----TACAAGAACCTGAGAGAG-----AGA 1296  
Db 323 SerValCysGluArgProHisTyrLysAsnLeuPheGluAspGlnAsnThrLeuThrSer 342  
QY 1297 ATCTGCAAGAAATATCTCTATGGGATCTGTATTGGAGA-----GAGGCAAA 1344  
Db 343 IleCysGluLysValIle-ValProAsnMetGluPheArgAlaAlaAspGluGluAla-P 362  
QY 1345 GACAGACTCAGAAACCTCTATGTTCCACAGCATTCCTTCTTTCAGCAGCTGAACGA 1404  
Db 362 heGluAspAsnSerGluGluTyIleArgAspLeu----- 374  
QY 1405 AGAACCATCAGATTTGGAGACCTCAACGAGGTTTCCTGAGAAATCTGTTCCGTTCTA 1464  
Db 375 --GluGlySerAspIleAspThrArgArgAlaAlaCysAspLeuValArgGlyLeuC 394  
QY 1465 CGGAATGGATGTTCCAAAACCTCTCTACCCCTTACCGAAAGATTCT---GGAGAAACAT 1521  
Db 394 yLysPhe-----PheGluGlyProValThrGlyIlePheSerGlyTyVala 410  
QY 1522 CTCTCT-----AGCTATACGAAG-----TGGCTTGACAGAGTCAG 1557  
Db 410 snSerMetLeuGlnGluTyAlaLysAsnProSerValAsnTrpLysHisLysAspAlaA 430  
QY 1558 GTATCAACTGGTTTATAATGCGCGGTTAGGCTATGGAAGCCGATGATAGAAGAGATG 1617  
Db 430 laiIeTyLeuValThrSerLeuAla----- 438  
QY 1618 TATCGGGAATGAACAGAGAGAGCTCAATAACAGTCACATTCACCCACCGAACCAAC 1676  
Db 439 -----SerLysAlaGlnThrGlnLysHisGlyIleThrGlnAlaAsn 452

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 2, 2004, 12:11:49 : Search time 332.5 Seconds  
(without alignments)  
3574.213 Million cell updates/sec

Title: US-09-423-575-1  
Perfect score: 3284  
Sequence: 1 attattgtataaaattgtgc.....ttaaaaaaaatataaaaaa 1886

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1298764 seqs, 315065143 residues

Total number of hits satisfying chosen parameters: 2597529

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09423575/runat\_01092004.161109.18704/app.query.fasta.1.2055  
-DB=Published Applications\_AA -QMT=fastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09423575@cgn.1.161@runat\_01092004.161109.18704  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:\*  
1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/prodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/prodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/prodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/prodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/prodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/prodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/prodata/1/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/prodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/prodata/1/pubpaa/US09C\_NEW\_PUB.pep:\*  
12: /cgn2\_6/prodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/prodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/prodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/prodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/prodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/prodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

ALIGNMENTS

RESULT 1  
US-10-424-599-273470  
; Sequence 273470, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 273470  
; LENGTH: 702  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MFT3847\_88966C.1.pep  
US-10-424-599-273470  
Alignment Scores:

	1	503	15.3	702		
Sequence 273470, A	12	15.3	702	12	US-10-424-599-273470	
Sequence 37053, A	12	473.5	14.4	550	12	US-10-425-114-37053
Sequence 336, App	3	472	14.4	680	12	US-10-425-066A-336
Sequence 2672, App	4	472	14.4	680	15	US-10-374-780A-2672
Sequence 56821, A	5	467.5	14.2	475	12	US-10-425-114-56821
Sequence 68805, A	6	467.5	14.2	499	12	US-10-425-114-68805
Sequence 122156, A	7	467.5	14.2	594	16	US-10-437-963-122156
Sequence 110193, A	8	465	14.2	618	16	US-10-437-963-110193
Sequence 162637, A	9	454	13.8	678	16	US-10-437-963-162637
Sequence 180416, A	10	443	13.5	538	12	US-10-424-599-180416
Sequence 43080, A	11	440.5	13.4	481	12	US-10-425-114-43080
Sequence 134068, A	12	439	13.4	642	16	US-10-437-963-134068
Sequence 48619, A	13	437	13.3	335	12	US-10-425-114-48619
Sequence 143490, A	14	431.5	13.1	640	16	US-10-437-963-143490
Sequence 193114, A	15	438.5	13.0	133	12	US-10-424-599-193114
Sequence 41924, A	16	428	13.0	447	12	US-10-425-114-41924
Sequence 208750, A	17	427	13.0	447	12	US-10-424-599-208750
Sequence 170412, A	18	426.5	13.0	884	16	US-10-437-963-170412
Sequence 268, App	19	424	12.9	739	12	US-10-425-068-268
Sequence 236, App	20	424	12.9	739	15	US-10-425-068-236
Sequence 2330, App	21	424	12.9	739	15	US-10-374-780A-2330
Sequence 42253, A	22	422	12.9	444	12	US-10-425-114-42253
Sequence 198122, A	23	422	12.9	463	12	US-10-424-599-198122
Sequence 40365, A	24	422	12.9	475	12	US-10-425-114-40365
Sequence 276400, A	25	421.5	12.8	461	12	US-10-424-599-276400
Sequence 117364, A	26	418	12.7	1123	16	US-10-437-963-117364
Sequence 39647, A	27	415.5	12.7	413	12	US-10-412-699B-39647
Sequence 270, App	28	412.5	12.6	610	15	US-10-412-699B-270
Sequence 26716, A	29	412.5	12.6	610	15	US-10-225-068-56
Sequence 206716, A	30	412.5	12.6	610	15	US-10-374-780A-2332
Sequence 37177, A	31	411.5	12.5	442	12	US-10-425-114-37177
Sequence 638, App	32	410.5	12.5	627	12	US-10-412-699B-638
Sequence 820, App	33	408	12.4	627	12	US-10-425-066A-820
Sequence 2482, App	34	408	12.4	627	14	US-10-286-264-18
Sequence 282020, A	35	408	12.4	627	15	US-10-374-780A-2482
Sequence 108604, A	36	407.5	12.4	654	12	US-10-424-599-282020
Sequence 57007, A	37	407.5	11.8	601	16	US-10-437-963-108604
Sequence 137941, A	38	386.5	11.7	377	12	US-10-425-114-57007
Sequence 3793, A	39	384	11.7	603	16	US-10-437-963-137941
Sequence 284132, A	40	383.5	11.5	224	12	US-10-425-114-37932
Sequence 47120, A	41	379	11.5	374	12	US-10-424-599-284132
Sequence 284018, A	42	378	11.5	374	12	US-10-425-114-47120
Sequence 39635, A	43	378	11.3	479	12	US-10-424-599-284018
	44	370	11.3	385	12	US-10-425-114-39635
	45	359.5	10.9	385	12	US-10-425-114-39635

```

QY 1324 TCTGTATTGAGAGAGGCAAAAGC-----AAGACTCAAGAAAACCTCTATG 1366
Db 339 LysThrLeu-----GlyGluAspAspCysLeuGlyValLeuGlySerArgLeu 356
QY 1369 -----TTCACACGATTCCTTCCAGAGCTGAAACGAAAGAAC 1410
Db 357 ArgTyrValAspHisHisLeuArgGlnArgAlaLeuGlnLeuGlyMetIleGln 376
QY 1411 CATCAGATTGGAGACCTCAACAGAGTTTCCTCGAGAAATCTGTTTCGGTTCACGGAAT 1470
Db 377 ProAsnAlaTrpArgProGlnArgGlyLeuProGluArgAlaValSerIleLeuArgAla 396
QY 1471 TGAATGTCCAAACACTTCCTTCCACCTTACCCGAAAGATTCGGAGAAACATCTTCAGCT 1530
Db 397 TrpLeuPheGluHisPheLeuHisProTyrProLysAspSerAspLysValMetLeuAla 416
QY 1531 ATACGAAGTGGCTTGACAAAGCTCAGGTATCAAACTGGTTTATAATGCGCGGGTTAGG 1590
Db 417 LysGlnThrGlyLeuAlaArgSerGlnValSerAsnTrpPheIleAsnAlaLeuValArg 436
QY 1591 CTATGGAAGCCGATGATAGAGAGATGTATGCGGNAATGAACAAGAGGAGCTCAATAAC 1650
Db 437 LeuTrpLysProMetValGluGluMetTyrLeuGluGluLeuLys----- 451
QY 1651 AGTCACATTCAACCCCAACCGA 1671
Db 452 GluHisGluGlnGlyAsnGly 458

RESULT 2
US-10-425-114-37053
; Sequence 37053, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 37053
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3060-003-B11_FLI_1.pep
US-10-425-114-37053

Alignment Scores:
Pred. No.: 1,7e-37 Length: 550
Score: 473.50 Matches: 128
Percent Similarity: 47.50% Conservative: 62
Best Local Similarity: 32.00% Mismatches: 119
Query Match: 14.42% Indels: 91
DB: 12 Gaps: 13

US-09-423-575-1 (1-1886) x US-10-425-114-37053 (1-550)
QY 550 GTTAGTTCATGGCTCTCCCTCTCTCATCTCTCCACTTCATCTTCGATCATTTAAGA 609
Db 85 LeuGlnValGlnAlaProPro----- 92
QY 610 CACTATGATGATTCCTCAACCAACATGG----- 639
Db 93 -----SerGlyTyrAsnLeuTrpThrProAlaMetAlaGlyAlaAlaAsp 108
QY 640 ---GGTTTGAGCAAAATGATGAGTTTCAGGCATTTTCAGGTGTAGTTGGTCCAGTGAA 696
; ; ; ; ;
; ; ; ; ;
; ; ; ; ;

```



Db 109 MetSerProGlnAlaGlnThrProGlyAlaAlaGlyAlaValSerAla 128  
Qy 697 CCAATGATGCTACATTCGGTGAAGAAGATTTCCCGTTTCTTAATTCGAATAAAGAAAC 756  
Db 128 -----  
Qy 757 AATGAGCTTTTCATTCAGTCTTGCATCAGATGTTTCTGATGAATGCTCGGAGATAAGTCTT 816  
Db 129 ---ValLeuSerLeuSer-----SerArgGluAlaProProValThrVal 142  
Qy 817 TGTGAGCTACAGATTAGCTCCTCAGAGCAAGTCTTCTGAGC-----AGCAAGACAT 870  
Db 143 AlaAlaVal-----AlaAlaCysThrAspGluGlyLysThrLeu 155  
Qy 871 TCTAATAAGCTTCTTACTCAAGGTTCTCTCAACTATATATTTGGCTCAAAATACCTTCA 930  
Db 156 GlyValSerAlaThrProGlnGly-----GlnMetValMetSerSerLysThrLys 173  
Qy 931 TCTGTTCAAGAAATACATTCATTCATTCGCGGATACCTCGCTCGATATTATTCATCTCGAGA 990  
Db 174 AlaAlaGlnGluLeuLeuAspGluValValSerLysGlyValGluAlaLys 193  
Qy 991 ACGAGTCAGGAGCTGCTAGTTCAGCCTTTACTTCA---CGTTTTCAGATATAACTGAG 1047  
Db 194 ThrAlaAlaAlaAlaThrLysSerLeuAlaAlaValLysLysGluAspSerGlu 213  
Qy 1048 TTTCTTTCAGTGGT-----GATTCTTAATACTCGAGCGGTTTCGGA----- 1089  
Db 214 GlyAlaSerGlyGlyGlyThrAspAspGlyAlaGlyAlaLysSerGlyGlyAlaAla 233  
Qy 1090 -----TCTACATTTCAAGGAGGAGCATTTAGAGCAAGAAACCCATCTCTTGAT 1140  
Db 234 ProGluMetSerThrAlaGluArgGlnGluMetGlnMetLysLysSerLysLeuLeuAsn 253  
Qy 1141 CTTCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
Db 254 MetLeuAspGluValGluGlnArgTyrArgGlnTyrHisGlyGlnMetGlnAlaValSer 273  
Qy 1201 TCAGCGTTCCTGCTGCAACCGAGTTA---GATCCACAGTTATACACACCGGTTCCCTC 1257  
Db 274 SerSerPheGluAlaAlaAlaGlyAlaGlySerAlaArgThrTyrThrAlaLeuAlaLeu 293  
Qy 1258 CAACCGTTTCTCTTATACAGACCTGAGAGAGAGATCTGCAAGAGATATCTCT 1317  
Db 294 ArgThrIleSerArgGlnPheArgCysLeuArgAspAlaIleAlaSerGlnValArgAla 313  
Qy 1318 ATGGGATCTGTATGAG 1365  
Db 314 AlaSerArgAlaLeu-----GlyGluAspAlaAspAlaAlaValAlaGlyArgThr 331  
Qy 1366 -----ATGTTCCACAGCATTCCTCTTTCAGCAG 1395  
Db 332 ValGlySerArgLeuArgTyrIleAspHisGlnLeuArgGlnArgAlaLeuGlnGln 351  
Qy 1396 CTGAACGAAAGAACCATCAGATTTCGAGACCTCAACGAGGTTTGCTGAGAAATCTGTT 1455  
Db 352 LeuGlyMetMetGlnGlyAlaThrArgProGlnArgGlyLeuProGluArgSerVal 371  
Qy 1456 TCGGTTCTACGAAATGAGATGCTCAAACTTCCTTCACCTTACCCGAAAGATTCGAG 1515  
Db 372 SerIleLeuArgAlaThrLeuPheGluHisPheLeuHisProTyrProLysAspSerAsp 391  
Qy 1516 AATCATCTCTAGCTATACGAGTGGCTGACAGAGAGTCAGGTATCAACTGGTTTATA 1575  
Db 392 LysIleMetLeuAlaLysGlnThrGlyLeuThrArgSerGlnValSerAsnThrPheIle 411  
Qy 1576 AATGCGCGGTTAGGCTTGAAGCCGATGATGAGAGAGATGATGATGAGAGATGATGAGAGAG 1635  
Db 412 AsnAlaArgValArgLeuThrLysProMetValGluGluMetTyrLeuGluThrLys 431

RESULT 3

US-10-225-066A-336  
; Sequence 336, Application US/10225066A

Publication No. US20030226173A1  
GENERAL INFORMATION:  
; APPLICANT: Mendel Biotechnology, Inc.  
; APPLICANT: RATCLIFFE, Oliver  
; APPLICANT: RIECHMANN, Jose Luis  
; APPLICANT: ADAM, Luc J  
; APPLICANT: DUBELL, Arnold T  
; APPLICANT: HEARD, Jacqueline E  
; APPLICANT: PILGRIM, Marsha L  
; APPLICANT: JIANG, Cai-Zhong  
; APPLICANT: REUBER, T. Lynne  
; APPLICANT: GREENMAN, Robert A  
; APPLICANT: PINEDA, Omaira  
; APPLICANT: YU, Guo-Liang  
; APPLICANT: BROUN, Pierre E  
; TITLE OF INVENTION: Yield-Related Polynucleotides and Polypeptides in Plants  
; FILE REFERENCE: MEI0036-2 US  
; CURRENT APPLICATION NUMBER: US/10/225,066A  
; CURRENT FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: 09/837,444  
; PRIOR FILING DATE: 2001-04-18  
; PRIOR APPLICATION NUMBER: 60/310,847  
; PRIOR FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: 60/336,049  
; PRIOR FILING DATE: 2001-12-05  
; PRIOR APPLICATION NUMBER: 60/338,692  
; PRIOR FILING DATE: 2001-12-11  
; PRIOR APPLICATION NUMBER: 10/171,468  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 1122  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 336  
; LENGTH: 680  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-10-225-066A-336

Alignment Scores:  
Pred. No.: 2,61e-37 Length: 680  
Score: 472.00 Matches: 152  
Percent Similarity: 42.01% Conservative: 74  
Best Local Similarity: 28.25% Mismatches: 193  
Query Match: 14.37% Indels: 119  
DB: 13 Gaps: 13  
US-09-423-575-1 (1-1886) x US-10-225-066A-336 (1-680)  
Qy 283 ATAAAGTTTCATAGAAACCCCAATGGACACACAAACAAACAACTTTTAGTTCTCTG 342  
Db 32 ValGlnTyrThrGlnGlnAspAspSerAsnAsnAsnAsnSerAsnSerAsn 51  
Qy 343 GAT-----AATGTCATGACTACCAAAATCCTCTTCTCATGGATTTTATACCTCA 393  
Db 52 AsnAsnAsnThrAsnThrAsnThrAsnAsnAsnSerSerPheValPheLeuAspSer 71  
Qy 394 AGAGAAGATTCAACTTCATCTCAACAATGTTCCATGGAAATACCATCAGATCAGATCCT 453  
Db 72 HisAlaPro-GlnProAsnAlaSerGlnGlnPheValGlyLeuProLeu----- 87  
Qy 454 CTACAAATGGGTGGCTTTGATATTTTCAATTCATGCTGACTACAAATATCTATCTATCT 513  
Db 88 -----SerGlyHis-- 90  
Qy 514 TCTCCAGGTCATCGATGTTCAAGATAACCGAATGTTGAGTTCATGGCTCTCTCTCTCT 573  
Db 91 GluAlaAlaSerIleThrAlaAlaAspAsnIleSerValLeuHisGlyTyrProProArg 110  
Qy 574 CATCCTCTCCATCTCATCTCTTTGGATCATTAAAGACACTGATGATGATGATGATGATGAT 633  
Db 111 ValGlnTyrSerLeuTyrGly-----SerHisGlnValAspProThrHisGln 126  
Qy 634 ATGTTGGGTTTGAAGCAATAGTAGTTTCAGGATTTTCAGGTTGAGTTGGTCCAGT 693



Db 32 ValGlnTyrThrGlnGlnAspSerAsnAsnAsnAsnAsnSerAsnAsnSerAsn 51  
QY 343 GAT-----AATGTCATGACTAACCAAAATCTCTTCTCATGGATTTTATACCTTCA 393  
Db 52 AsnAsnAsnThrAsnThrAsnThrAsnAsnAsnAsnSerSerPheValPheLeuAspSer 71  
QY 394 AGAGAAGATTCAACTTCTTCAACAATGCTTCCATGCAATACCATCAGATCAGATCCT 453  
Db 72 HisAlaPro-GlnProAsnAlaSerGlnGlnPheValGlyIleProLeu----- 87  
QY 454 CTCAAAATGGTGGCTTTGATATTTTCAATTTATGCTGACTAACAAATACTATCATCT 513  
Db 88 -----SerGlyHis-- 90  
QY 514 TCTCCAGGCTCATGATGTTCAAGATAACCCCAATGTTGATTCATGGCTCTCTCTCTCT 573  
Db 91 GluAlaAlaSerIleThrAlaAlaAspAsnIleSerValLeuHisGlyTyrProProArg 110  
QY 574 CATCTCTCCCTCACTTCTCATCTTGGATCATTTTAAGACACTATGATGATTCCTCAACAAC 633  
Db 111 ValGlnTyrSerLeuTyrCly-----SerHisGlnValAspProThrHisGln 126  
QY 634 ATGTGGGGTTTGAAGCAAAATAGTGAAGTTTCAGGCATTTTCAGGTGTAGTTGGTCCAAAGT 693  
Db 127 GlnAlaAlaCysGluThrProArgAlaGlnGlnGlyLeuSerLeuThrLeuSerSerGln 146  
QY 694 GAACCAATGATGCTACATTCGGTGAAGAAGATTTCCCGTTTCTTAATTCGAATAAAGA 753  
Db 147 GlnGlnGlnGlnGlnHisGlnHisGlnProIleHisValGlyPheGlySer 166  
QY 754 AACAAATGAGCTTTCATGAGTTCATGCTGATGCTGATGTTTCTGATGAATGCTCGAGATAAGT 813  
Db 167 GlyHisGlyGluAspIleArgValGlySer----- 176  
QY 814 CTTTGTGAGCTACAGATTACCTTCAGAGCAGCTTCTTCGAGCAGCAAGACATTTCT 873  
Db 177 -----GlySerThr 179  
QY 874 AATAACGTTGTTACTCAAGTTTCTTCAACTTATATTTGGCTCAAAATACCTTCACTCT 933  
Db 180 GlySerGlyValThrAsnGlyIleAlaAsnLeuVal---SerSerLysTyrLeuLysAla 198  
QY 934 GTTCAAGAAATACTATCTCATTTCCGCGCATCTCGCTCGATATTATCTATCTCGAGAAC 993  
Db 199 AlaGlnGluLeuLeuAspGluValValAsnAlaAspSerAsp----- 212  
QY 994 GAGTCAGAGCTGCTAGTTCACGCTTCTTACTTCACTGTTTGGAGAATA---ACTGAGTTT 1050  
Db 213 AspMetAsnAlaLysSerGlnLeuPheSerSerLysLysGlySerCysGlyAsnAspLys 232  
QY 1051 CTTGATGCTGATTTCTAATAACTCGGAGCGGGTTTCCGATCT----- 1092  
Db 233 ProValGlyGluSerSerAlaGlyAlaGlyGlyGlyGlySerGlyGlyValaGluAla 252  
QY 1093 -----ACATTTCAAGGAGAGCATTAGAAGCAAGAA 1125  
Db 253 AlaGlyLysArgProValGluLeuGlyThrAlaGluArgGlnGluIleGlnMetLysLys 272  
QY 1126 ACCCATCTCTGATCTTCTTCAATGTTGATGATGATGATGATGATGATGATGATGATGAT 1185  
Db 273 AlaLysLeuSerAsnMetLeuHisGluValGluGlnArgTyrArgGlnTyrHisGlnGln 292  
QY 1186 ATTCATACGTTTATATCAGCTTCCATGCTGCAACCGAGTGA---GATCCAGATTACAC 1242  
Db 293 MetGlnMetValIleSerSerPheGluGlnAlaAlaGlyIleGlySerAlaLysSerTyr 312  
QY 1243 ACCCGGTTTGGCTCCAAACCGTTTCTTCTTATACAAAGACCTCGAGAGAGATCTGC 1302  
Db 313 ThrSerLeuAlaLeuLysThrIleSerArgGlnPheArgCysLeuLysGluAlaIleAla 332  
QY 1303 AAGAAGATAATCTCTATGCGATCTCTATTTGGAGAGAGGCAAGAACACAGACTCAAGAAC 1362  
Db 333 -----GlyGlnIleLysAlaAlaAsnLysSerLeuGlyGluGluAsp 346

QY 1363 TCTATG-----TTCACCCAGCATTTGCTT 1386  
Db 347 SerValSerGlyValGlyArgPheGluGlySerArgLeuLysPheValAspHisHisLeu 366  
QY 1387 CTTTCAGCAG-----CTGAACGAAAGAACCATCAGATTGG 1422  
Db 367 ArgGlnGlnArgAlaLeuGlnGlnLeuGlyMetIleGlnHisProSerAsnAlaIleP 366  
QY 1423 AGACCTCAACGAGGTTTGGCTTGAGAAATCTGTTTCGGTTCTACGGAATTGGATGTTCCAA 1482  
Db 387 ArgProGlnArgGlyLeuProGluArgAlaValSerValLeuArgAlaIlePLeuPheGlu 406  
QY 1483 AACTTCCTTCACTTACCCGAAAGATTCGAGAAACATCTTCTAGCTATACGAATGGC 1542  
Db 407 HisPheLeuHisProTyrProLysAspSerAspLysHisMetLeuAlaLysGlnThrGly 426  
QY 1543 TTGACAAAGAGTCAAGTATCAAACTGTTTATATAATGCGCGGTTAGGCTATGCAAGCCG 1602  
Db 427 LeuThrArgSerGlnValSerAsnTyrPheIleAsnAlaArgValArgLeuTyrLysPro 446  
QY 1603 ATGATAGAAGAGATGATGCGGAAATGCAACAGAGAGCTCAATTAACATCATTCATCAA 1662  
Db 447 MetValGluGluMetTyrMetGluMetLysGluGlnAlaLysAsnMetGlySerMet 466  
QY 1663 CCAACCGACCA-----ACTCTTCGAATG 1686  
Db 467 GluLysThrProLeuAspGlnSerAsnGluAspSerAlaSerLysSerThrSerAsnGln 486  
QY 1687 CCAAAATCTGTTATGATGAGCCCAAGCAATGCATAAATAAGACAAACAAT 1734  
Db 487 GluLysSerProMetAlaAspThrAsnTyrHisMetAsnProAsnHis 502  
RESULT 5  
US-10-425-114-56821  
; Sequence 56821, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 56821  
; LENGTH: 475  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17364H11\_FLI.pep  
US-10-425-114-56821  
Alignment Scores:  
Pred. No.: 6,27e-37 Length: 475  
Score: 467.50 Matches: 144  
Percent Similarity: 46.67% Conservative: 80  
Best Local Similarity: 30.00% Mismatches: 175  
Query Match: 14.24% Indels: 81  
DB: 12 Gaps: 14  
US-09-423-575-1 (1-1886) x US-10-425-114-56821 (1-475)  
QY 292 CATGAAGAACCATGAGACACACACACACACACTTTTCTGATATGTC 351  
Db 2 HisGlnSerAspSerAspAsnAsnSerHisTyrLeuProSerSerLeuHisArgSer 21  
QY 352 ATGACTAACCAAAATCTCTTCTCTGATTTTATACCTTCAAGAGAGATTCACATTC 411



QY 1054 GATGGT-----GATTCTTAATACTCGAGGCGGGTTTCGGA----- 1089  
Db 165 SerGlyGlyThrAspAspGlyAlaGlyAlaLysSerGlyGlyGlyAlaAlaProGlu 184  
QY 1090 ---TCTACATTTCAAGAGAGACATTAGAACGCAAGAAACCCATCTCTTGATCTTCT 1146  
Db 185 MetSerThrAlaGluArgGlnMetGlnMetLysSerLysSerLysLeuAsnMetLeu 204  
QY 1147 CAATGGTGGATGATCATATAGTCGCTGATGAGATTCAATACGTTTATATACAGC 1206  
Db 205 AspGluValGluGlnArgTyrArgGlnTyrHisGlyGlnMetGlnAlaValSerSer 224  
QY 1207 TTCCATGCTGCACCGAGTTA---GATCCACAGTTACACACCGGTTGCTCCCAACCC 1263  
Db 225 PheGluAlaAlaGlyAlaGlySerAlaArgThrTyrThrAlaLeuAlaLeuArgThr 244  
QY 1264 GTTTCCTCTTATACAAGAACCTGACAGAGAGAACTGCAAGAGAAATCTCTATGGGA 1323  
Db 245 IleSerArgGlnPheArgCysLeuArgAspAlaIleAlaSerGlnValArgAlaAlaSer 264  
QY 1324 TCTGATTGGAGAGGCAAGACACAGACTCAAGAAACCTCT----- 1365  
Db 265 ArgAlaLeu-----GlyGluAspAlaAspAlaAlaValAlaGlyGlyArgThrValGly 282  
QY 1366 -----ATGTTCCACGACGATTCCTCTCTTCAGCAGCTGAAA 1401  
Db 283 SerArgLeuArgTyrIleAspHisGlnLeuArgGlnArgAlaLeuGlnLeuGly 302  
QY 1402 CGAAGAACCATCAGATTTGGAGACCTCAACGAGGTTTCCTGAGAAATCTGTTTCGTT 1461  
Db 303 MetMetGlnGlyGlyAlaTrpArgProGlnArgGlyLeuProGluArgSerValSerile 322  
QY 1462 CTACGGAATTTGATGTTCCAAACTCTCTTCCACCTTACCCGAAAGATTCGAGAGAAAT 1521  
Db 323 LeuArgAlaTrpLeuPheGlnHisPheLeuHisProTyrProLysAspSerAspLysile 342  
QY 1522 CTTTACGTATACGAAGTGGCTTGACAGAAGTCAGGTATCAAACTGGTTTATAAATCGG 1581  
Db 343 MetLeuAlaLysGlnThrGlyLeuThrArgSerGlnValSerAsnTrpPheIleAsnAla 362  
QY 1582 CGGTTAGCTATGAGACCGGATGATAGAGAGATGTATCGGAATGACAAAG 1635  
Db 363 ArgValArgLeuTrpLysProMetValGluGluMetTyrLeuGluGluThrLys 380

RESULT 7  
US-10-437-963-122156  
; Sequence 122156, Application US/10437963  
; Publication No. US2004012343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 122156  
; LENGTH: 594  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_2510C.1.p  
US-10-437-963-122156  
Alignment Scores:  
Pred. No.: 6.89e-37 Length: 594

Score: 467.50 Matches: 123  
Percent Similarity: 45.90% Conservatives: 56  
Best Local Similarity: 31.54% Mismatches: 98  
Query Match: 14.24% Indels: 113  
DB: 16 Caps: 10  
US-09-423-575-1 (1-1886) x US-10-437-963-122156 (1-594)  
QY 757 AATGAGTTCATTGATGCTTTCATCATGATGTTTCT----- 792  
Db 182 SerGluLeuSerLeuSerLysSerLysSerSerAspSerMetLeuAsnAlaGly 201  
QY 793 ---GATCAATCTCGGAGATAAGTCTTTGTCAGCTACAAAGATTAGCCTCAGAGCAAGCT 849  
Db 202 GlyAspGlnCysSerSer-----AlaAlaSerArgSerGlyLeuThrGln--- 216  
QY 850 TCTTGCAGACGAAAGACATTTCTTAATACGTTGTTTACTCAA----- 891  
Db 217 ---MetSerArgValValValGluProGluProGluProLeuVal 230  
QY 892 -----GTTTCTCTCAACTTATATTTGGCTCAAAATACCTTCACTCT 933  
Db 231 ProTyrTyrProAlaAlaAsnPheAlaValValAlaArgSerArgTyrAlaAlaVal 250  
QY 934 GTTCAAGAAATATCTATCTCATTTCT----- 957  
Db 251 AlaGlnGlnValLeuAsnAspAlaValGlyCysValLeuGlyValAlaAspAlaAla 270  
QY 958 -----CCGCATACTCGCTCGATTAT---TCATCTCGAGAACCGAGCTCAGAGCTGCT 1008  
Db 271 AlaAspSerAlaSerGlyValAspSerGlySerSerArgProSerSerCysSerValAla 290  
QY 1009 AGTTCAGCCTTTACTTCACGCTTTTGAATATAAATCACTGAGTTTCTGTGATGTTCTAAT 1068  
Db 291 GlyGlyAlaProSerSerAlaValSerSerAsnAsnGlnLeuIleAlaSerSerGlyGlu 310  
QY 1069 AACTCGAGGCGGGTTTCGAGATCTACATTTCAAAGGAGACATTAGAACGAAAGAAACC 1128  
Db 311 HisThrHisGlyGlyGlyAspAlaSerAlaGlnArg-----LeuArgSer 325  
QY 1129 CATCTCTTGGATCTTCTTCAAAATGGTG----- 1155  
Db 326 GluLeuLeuThrMetLeuGlnLeuValSerSerProPheSerThrArgLeuProArgHis 345  
QY 1155 ----- 1155  
Db 346 GlnPheLysGlnIleSerTyrTyrHisPheGluSerAsnLeuLysProSerPheAspAsp 365  
QY 1156 -----GATCATCGATATAGTATGATTCCTGCTAGATGATTCAT 1191  
Db 366 AspThrAsnSerAsnAlaGlnMetAspGlnLysTyrAsnGlnCysLeuAspGluIleGln 385  
QY 1192 ACGGTTATATCAGCGTTCCATGCTGCAACCGAGTTA----- 1227  
Db 386 SerThrThrAlaArgPheAsnThrLeuThrHisAlaThrAlaArgAlaAlaGlyMetSer 405  
QY 1228 GATCCACAGTTACACCCCGTTTGCCTCCAAACCGTTTCTTCTTATACAGAACCTG 1287  
Db 406 SerSerSerIleCysAlaProPheAlaHisArgAlaValSerAlaMetTyrHisGlyLeu 425  
QY 1288 AGACGAGAAATCTCAAGAGAGATAATCTCTATGGATCTGTATTGGAGAGA----- 1338  
Db 426 ArgArgGlnAlaGlyGluLeuMetSerAlaAlaAlaAlaAlaGlyArgProCysArg 445  
QY 1339 -----GGCAAGACAGACTCAAGAAACCTCT 1365  
Db 446 GlyGlyGluSerSerSerAlaValThrGlyGlyGluArgGluArgSerTrpGluSerAla 465  
QY 1366 ATGTTCCACGAGCATTCCTCTTTCAGCAGCTGAACGAAAGAACCATTCAGATTGAGGA 1425  
Db 466 PheIleGlnLysHisTrpAlaValGlnGlnLeuArgGlyGluGlnGlnCysTrpArg 485  
QY 1426 CCTCAACGAGGTTTGCCTGAGAAATCTGTTTCGGTTCTACGGAATGATGTTCCAAAC 1485

Db 486 ProGlnArgGlyLeuProGluLysSerValAlaValLeuLysAlaTrpMetPheGluAsn 505  
 QY 1486 TTCCTTCAACCTTACCGAAGATTCGAGAACATCTTCTAGCTATACGAGTGGCTTG 1545  
 Db 506 PheLeuArgProTyrProLysAspSerGluLysGluMetLeuAlaAlaArgSerGlyLeu 525  
 QY 1546 ACAAGAGTCAGTATCAACCTGGTTTATAAATGCGGGTTAGCTATGGAAGCCGATG 1605  
 Db 526 SerArgAsnGlnValSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysPromet 545  
 QY 1606 ATAGAAGAGATGATCGCGAAATGAACAAG 1635  
 Db 546 IleGluGluMetCysGluGluLeuLysArg 555

## RESULT 8

US-10-437-963-110193  
 ; Sequence 110193, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53221)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 110193  
 ; LENGTH: 618  
 ; TYPE: PRT  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_1427C.1.pap  
 US-10-437-963-110193

Alignment Scores:  
 Pred. No.: 1,248-36 Length: 618  
 Score: 485.00 Matches: 141  
 Percent Similarity: 46.53% Conservative: 60  
 Best Local Similarity: 32.64% Mismatches: 119  
 Query Match: 14.16% Indels: 112  
 DB: 16 Gaps: 16

US-09-423-575-1 (1-1886) x US-10-437-963-110193 (1-618)

QY 583 CCATTCATCTCTTGGATCAATTAAGACATATGATGATCTCTCAACAAACATGTGGGT 642  
 Db 147 ProLeuHisGlyAlaTyrTyrLeuSerGlyPhe-----SerSerGlyAlaGlyAlaGly 164  
 QY 643 TTGGAAGCAATAGTGGTTCAGGCATTTTCAGGTGTAGTGGTCCAAAGTGAACCAATG 702  
 Db 165 Phe-----LeuSerProPheAlaAlaSer-----ValAlaAlaAlaAlaProAla 180  
 QY 703 ATGTCTACATTCGGTGAAGAGATTTCCCGTTCTTAATTTTCAATTAAGAAAGAAACATGAG 762  
 Db 181 AlaSer-----Glu 183  
 QY 763 CTTTCATTGAGTCTTGCA-----TCAGATGTTCTGATCAATGC 801  
 Db 184 LeuSerLeuArgLeuGlyAlaThrLysCysSerSerProSerSerMetAlaAsnAlaSer 203  
 QY 802 TCGGAGATAAGTCTTTGTGACGCTACAAGATAGCTCA-----GAG 843  
 Db 204 SerGluValSerCysSerGlyLeuThrHisValSerSerGlyGlyLeuGlyTyrHis 223  
 QY 844 CAACTTCTTGCAGCAGCAACACATTTCTTAATACGTGTGTTACTCAAGT----- 894

Db 224 GluAlaAlaAlaGlyAlaGlyAlaAlaLeuPheHisProThrHisGlyAspAspAla 243  
 QY 895 -----TTC 897  
 Db 244 AlaAlaAlaAlaAlaGlyGluLeuArgGlnAlaTyrHisSerArgAlaProProHisPhe 263  
 QY 898 TCTCACTTATATTGGCTCAAAATACCTT---CACTCTGTTCACAGAAATACATCTCAT 954  
 Db 264 SerGlnValValSerArgSerAlaValLeuAlaHisValAlaGlnGluLeuLeuAsnGly 283  
 QY 955 TTGCGCGCATACTCGCTCGAT-----TATTCTCTCGAGAAACCGAGTCAGGA 1002  
 Db 284 PheValAlaCysLeuLeuGlnAspValAlaAlaAspAlaAlaSerGlyValAspGlyGly 303  
 QY 1003 GCTGCTAGTTCAGCCTTTACTTCACGTTTGAGAAATATACTGAGTTTCTGATGCTGAT 1062  
 Db 304 GluAlaSerGlnAla-----LeuLeuValGlyLeuPheGlyGlu 316  
 QY 1063 TCTAATAACTCGGAGCGGGTTTCGGATCTACATTTCAAAGGAGA----- 1107  
 Db 317 AspHisaspGlyAlaAspGlyGlyCysLeuThrArgGluArgArgCysGluValGlyGly 336  
 QY 1108 -----GCATTAGAACAAAGAAACC----- 1128  
 Db 337 GlyGlyAlaAlaSerGluGluAlaThrProThrGlyGluPheLeuAlaGlnPheLeuVal 356  
 QY 1129 ---CATCTCTTGGATCTTCTT-----CAATGTGTGATGAT 1161  
 Db 357 GluHisLeuArgAsnArgPheLeuValArgThrAsnGlyAspGluMetGlnValaspGlu 376  
 QY 1162 CGATATAGTCATTCGTAGATGATCATACGGTTATATACGGTTTCCATGCTGCAACC 1221  
 Db 377 LysCysAsnGlnCysValGluGluMetGlnSerThrAlaAlaArgPheAsnSerMetVal 396  
 QY 1222 GAGTTA-----GATCCACAGTTACACACCGGTTTCCCTCCAAACCGTTCC 1269  
 Db 397 ArgSerThrGlyGlyGlyGlyLeuThrAlaAlaPheAlaGlyArgAlaValAla 416  
 QY 1270 TTCTTTACAAGAACCTGAGAGAGAAATCTCAAGAGATATATC----- 1314  
 Db 417 AlaAlaTyrArgArgValArgArgValMetGlyGlnLeuValAlaAlaAlaThrAla 436  
 QY 1315 -----TCTATGGATCTGTATTGGAGAGGCAAGACAGACTCAAGAA 1359  
 Db 437 ArgSerSerSerAlaAlaAlaAlaLeuGluGlu---LysGluArgSerTrpGlu 455  
 QY 1360 ACCTCTATGTTCCACCGAGATTCCTTCTTCAGCAGCTGAAACGAAAGAACCATCAGATT 1419  
 Db 456 SerSerPheIleGlnLysHisTrpAlaMetGlnGlnLeuArgArgGlyAspGlnGlnSer 475  
 QY 1420 TGAGACCTCAACGAGGTTTGGCTGAGAAATCTGTTTCGGTTCTACGGAATGGATGTTTC 1479  
 Db 476 TrpArgProGlnArgGlyLeuProGluLysSerValAlaValLeuLysAlaTrpMetPhe 495  
 QY 1480 CAAACTTCTCTCACCTTACCCGAAAGATTCCGAGAAACATCTTCTAGCTATACGAGT 1539  
 Db 496 GluAsnPheLeuArgProTyrProLysAspSerGluLysAspMetLeuAlaAlaArgSer 515  
 QY 1540 GGTTCGACAGAGTTCAGGTATCAAACTGGTTTATAAATGCGGGTTAGGCTATGGAAG 1599  
 Db 516 GlyLeuSerArgSerGlnValSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLys 535  
 QY 1600 CCGATCATAGAGAGATGTATCGGAAATGAACAAG 1635  
 Db 536 ProMetIleGluAspMetTyrGluGluLeuLysLys 547

## RESULT 9

US-10-437-963-162637  
 ; Sequence 162637, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.

APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 162637  
LENGTH: 678  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_6170C.1.p  
US-10-437-963-162637

Alignment Scores:  
Pred. No.: 1,59e-35 Length: 678  
Score: 454.00 Matches: 137  
Percent Similarity: 46.93% Conservative: 62  
Best Local Similarity: 32.31% Mismatches: 149  
Query Match: 13.82% Indels: 76  
DB: 16 Gaps: 14

US-09-423-575-1 (1-1886) x US-10-437-963-162637 (1-678)

QY 481 AATTCTATCTGACTACAAATACATATCATCT-----513  
DB 48 AsnMetMetLeuAsnSerAlaValThrSerMetThrProValSerPheGlyHisGln 67  
QY 514 ---TCCTCCAGGCTCTATCGATGTTCAAGATAACCGCAATGTTGAGTTCATGCTCCTCCT 570  
DB 68 ProSerProSerSerSerSerAlaGlnHis-----PheValGlyIlePro 83  
QY 571 CTTCTATCTCTCCACTTCATCTCTTGGATCATTTAAGACACTATGATGATCTCTCAAC 630  
DB 84 LeuGlnAlaProProAla-----SerGlyTyr 92  
QY 631 AACATGCTGGGTTTTCGAACAAATAGTGTGATTTTCAGGCATTTCAGTGTAGTGTGCTCA 690  
DB 93 AsnLeuThrProAlaAlaAlaThrGly-----AlaGlyAspMetSerPro 108  
QY 691 AGTGAACCAATGATGCTACATTCGCTGAAGAAGATTTCCCGTTCTTAATTTCGAATAAA 750  
DB 109 ProThrPro-----GlnHis 113  
QY 751 AGAACAATGAGCTTTTCATGAGTCTTCATCAGATGTTTCTGATGATGCTCGGAGATA 810  
DB 114 GlnHisGlnAlaHisGlyGlyAlaAlaGlyValSerAlaValLeuSerLeuSer 133  
QY 811 AGTCTTTGTGCGACT-----ACAAGATTAGCTCAGACCAAGCTTTCTTCGAGCAGC 861  
DB 134 SerArgGluAlaAlaProProValThrValAlaAlaValAlaAlaGlyAspGluGly 153  
QY 862 AAGACATTTCTTAATACGTTGTTTACTCAAGCT-----TTCTCTCACTTATATT 912  
DB 154 LysTyLeu-----GlnAlaValAlaGlnGlyAlaAlaSerHisGlyGlnMetValMet 171  
QY 913 GGTCTAAATACCTTCACCTCTGTTCAGAAATATCTATCTCAATTCGCGCATCTCGCTC 972  
DB 172 SerSerLysTyrLeuLysAlaAlaGlnGluLeuLeuAspGluValValSerValSerLys 191  
QY 973 GATTATTTCATCTCGAGAACCGAGTTCAGGCTGCTAGTTCAGCTTTTACTTACGCTTTT 1032  
DB 192 GlyValAspValLysAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaLys 211  
QY 1033 GAGAATATACTAGTTCCTTGTGCT-----CATTTCTAATACTCGGAGGG 1080

212 GluAsp---SerGluGlyValSerGlyGlyGlyThrGluAspGlyGlyGlyAlaLysSer 230  
QY 1081 GGTTCGGA-----TCTACATTTCAAGAGAGAGCATTAGAACAAAG 1122  
DB 231 GlyGlyAlaProProProGluMetSerThrAlaGluArgGlnGluLeuGlnMetLys 250  
QY 1123 AAAACCCATCTCTTGATCTTCTTAAATGCTGATCATCATATAGTCACTTCGCTAGAT 1182  
DB 251 LysGlyLysLeuLeuAsnMetLeuAspGluValGluGlnArgTyrArgGlnTyrHisGln 270  
QY 1183 GAGATTCATACGGTTATATCAGCGTTCATGCT---GCAACCGAGTTAGATCCACAGTTA 1239  
DB 271 GlnMetGlnValValValAlaSerPheGluAlaValAlaGlyGlyGlySerAlaArgThr 290  
QY 1240 CACACCGGTTTCCCTCCAAACCGTTTCTTATACAAACCTGAGAGAGAGATC 1299  
DB 291 TyrThrAlaLeuAlaLeuArgThrIleSerArgGlnPheArgCysLeuArgAspAlaIle 310  
QY 1300 TGCAGAAGATAATCTCTATGGGATCTGTATTG-----GAGAGAGGC 1341  
DB 311 AlaGlyGlnValArgAlaAlaSerArgAlaLeuGlyGluAlaValAspAlaAspGlyGly 330  
QY 1342 AAAGACAAGACTCAAGAAACCTCTATG-----TTCCACCAGCATTCG 1383  
DB 331 CysGlyArgThrValGlySerArgLeuArgTyrIleAspHisGlnLeuArgGlnArg 350  
QY 1384 CTTCTCAGCAGCTGAAACGAAAGACCATCAGATTTGGAGACTCAACGAGTTTCCT 1443  
DB 351 AlaLeuGlnGlnLeuGlyMetGlnSerSerAlaIleArgProGlnArgGlyLeuPro 370  
QY 1444 GAGAAATCTGTTTCGGTTCACGGAATTCGATGTTCCAAACTTCCTTCCCTTACCCG 1503  
DB 371 GluArgSerValSerIleLeuArgAlaThrLeuPheGluHisPheLeuHisProTyrPro 390  
QY 1504 AAGATTCGAGAAACATCTTCTAGCTATACGAGTGTGCTTACAAAGAGTTCAGTATCA 1563  
DB 391 LysAspSerAspLysIleMetLeuAlaLysGlnThrGlyLeuThrArgSerGlnValSer 410  
QY 1564 AACTGTTTATAATGCGGCTTAGCTTAGTGAAGCCGATGATAGAGAGATGTATCGC 1623  
DB 411 AsnIlePheLeuAlaAlaArgValArgLeuTrpLysProMetValGluGluMetTyrLeu 430  
QY 1624 GAAATGAACAAG 1635  
DB 431 GluGluThrLys 434

RESULT 10  
US-10-424-599-180416  
; Sequence 180416, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 180416  
; LENGTH: 538  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_133930C.1.p  
US-10-424-599-180416  
Alignment Scores:  
Pred. No.: 1.77e-34 Length: 538  
Score: 443.00 Matches: 133  
Percent Similarity: 46.67% Conservative: 77

Best Local Similarity: 29.56% Mismatches: 152  
Query Match: 13.49% Indels: 88  
DB: 12 Gaps: 17

US-09-423-575-1 (1-1886) x US-10-424-599-180416 (1-538)

QY 586 CTTTCATCCCTTGGATCAATTAAGACACATATGATATCC---TCAACAACATCGTGGGT 642  
Db 2 LeuAsnThrLeuHisLeuLeuLeuHisHisAspGlnArgHisLeuTrp---20  
QY 643 TTTGAAGCAAAATAGTGAGTTTCAGGCATTTTCAGGTGTAGTGTGCTCAAGTGAA-----696  
Db 21 -----PheGluLeuThrProSerValHisProProSerSerThrThr 34  
QY 697 -----CMAATGATG-----TCTACATTCGGTGAAGAA 723  
Db 35 ThrValProGlnIleThrLysThrThrGlnArgAlaThrAsnSerAlaLeuGlyGln---53  
QY 724 GATTTCCCGCTTTCTAATTTCCGAATAAAGAACAAATGAGCTTTTCATTGAGCTTTGTCATCA 783  
Db 54 -----ProCysLeuLeuSerLeuAlaLeuSerProLeuLeuAlaThrSerLeuLeuArg 71  
QY 784 GATGTTTCTGATGAATGCTCGGAGATAAGTCTT-----TGTGCGAGCTACAAGATTAGCC 837  
Db 72 GluPheGlySerAlaAlaTrpAlaArgHisProValTrpGlyValProValAlaArgLeuAla 91  
QY 838 TCAGACACAGCTTCITGCGAGCAGCAAGACACATTTCTAATAACGTGTGTACTCAAGTTTC 897  
Db 92 GlnGlnMetLeuSerGlnAspGlyTyrValSerSerAsnSerLysValAlaSerValTyr 111  
QY 898 TCTCAA-----CTTATATTTGGCTTCAAAATACCTTCACCTCTGTTCAAGAAATACTA 948  
Db 112 GlnGlnGlyHisPheLeuValLysAsnSerLysPheLeuValProAlaGlnAspLeuLeu 131  
QY 949 TCTCATTTCGCGCATCTCGCTCGATTAATTCATCTCGAGAACCGAGTCAGAGCTGCT 1008  
Db 132 AsnGluPheCysSer-----LeuCysAlaLysGlnSerAspLeuGlyLysPro 147  
QY 1009 AGTTCAGCGCTTTACTTCACGTTTTGAGAATAATACTGAGTTCTTGTGATGATCTAAT 1068  
Db 148 ThrLysSerLeuLysLysGlnTrpGlu-----AspGlnGlu 159  
QY 1069 ACTCGAGCGGGTTTCGGATCTACATTTCAAAGAGACATAGAA-----1116  
Db 160 AsnAsnGlyValGlySerSerLysLysHisSerLeuThrSerLeuGluPheValGluLeu 179  
QY 1117 ---GCAAAAGAAACCCATCTCTGGATCTCTTCAAATGGTGATGATCGATATAGTCAT 1173  
Db 180 GlnLysArgLysThrLysLeuLeuSerMetLeuGluLeuValAspArgTyrLysHis 199  
QY 1174 TCGGTAGATGATTCATACGGTTATATACGGTTCCATGCT---GCAACCGAGTTAGAT 1230  
Db 200 TyrArgAsnGlnMetLysSerValValSerSerPheGluAlaValAlaGlyAsnGlyAla 219  
QY 1231 CCACAGTTACACACCGGTTTGCCTCCAAACCGTTTCTTCTTATACAAAGACCTGAGA 1290  
Db 220 AlaThrValTyrSerAlaLeuAlaLeuLysAlaMetSerArgHisPheArgCysLeuLys 239  
QY 1291 GAGAGAAATCTCGAAGAAGATA-----ATC 1314  
Db 240 AspGlyIleLeuSerGlnIleGlnAlaThrArgLysAlaMetGlyGluLysAspProVal 259  
QY 1315 TCTATGGCATCTGATTGGAGAGGCGAAA-----GACAGACT 1353  
Db 260 AlaProGlyThrThr-----ArgGlyGluThrProArgLeuLysValIleAspGlnThr 277  
QY 1354 CAAGAAACCTCTATGTTCCACGACATGCTCTTCCAGCAGCTGGAACGAAAGAACCAT 1413  
Db 278 -----LeuArgGlnGlnArgAlaPheGlnGlnMetSerMetMetGluThr 292  
QY 1414 CAGATTTCGAGACCTCAACGAGGTTTCCCTCAGAAATCTGTTTCGGTTCTACGGAATGG 1473  
Db 293 HisProNtrpArgProGlnArgGlyLeuProGluArgAlaValSerValLeuArgAlaTrp 312

1474 ATGTTCCAAAATCTCTTACCTTACCGGAAAGATTCGGAGAAACATCTTCTTACGTATA 1533  
Indels: 88  
313 LeuPheGluHisPheLeuHisProTyrProSerAspValAspLysHisIleLeuAlaArg 332  
QY 1534 CGAAGTGGCTTGACAAAGTCAAGTATCAAACTGTTTATAAATCCGCGGTTAGGCTA 1593  
Indels: 88  
333 GlnThrGlyLeuSerArgGlyGlnValSerAsnTrpPheIleAsnAlaArgValArgLeu 352  
QY 1594 TGAAGCCGATGATACAGAGATGATGCGGAATATGAAACAAGAGGAACTCAATAAC---1650  
Indels: 88  
353 TrpLysProMetValGluGluMetTyrLeuGluValLysAspProGluAsnAlaIle 372  
QY 1651 -----AGTCACATTCACCAACCGGA-----1671  
Indels: 88  
373 AlaSerSerGluGlyAlaThrAspGlnAspAsnAspIleAsnProAsnAsnValGlnTyr 392  
QY 1672 ---CCAACCTCTCGAATGCCAAAATCTGTATGATGAGCAAGCAATGCATAAATAAGAC 1728  
Indels: 88  
393 ProProProLeuSerSerArgSerGluAspGlnLysProSerLeuValArgIleAsp 412  
QY 1729 AACAAATGTTTACCACACTTTGTGATAAT 1758  
Indels: 88  
413 SerGluCysAlaSerSerIleIleAsnAsn 422

RESULT 11  
US-10-425-114-43080  
; Sequence 43080, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingtong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 43080  
; LENGTH: 481  
; TYPE: PRT  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: 700684286\_FLI.pep  
US-10-425-114-43080

Alignment Scores:  
Pred. No.: 2,99e-34 Length: 481  
Score: 440.50 Matches: 97  
Percent Similarity: 63.43% Conservative: 40  
Best Local Similarity: 44.91% Mismatches: 60  
Query Match: 13.41% Indels: 19  
DB: 12 Gaps: 5

US-09-423-575-1 (1-1886) x US-10-425-114-43080 (1-481)

QY 1057 GGTGATCTTAATAACTCGGAGCGGTTTCGGA-----TCTACATTTCAA 1101  
Indels: 88  
6 GlyIleAsnAsnAsnAsnAspGlyGlyLysGlnGlyValGluLeuSerThrAlaGln 25  
QY 1102 AGGAGACATTAGAAGCAAGAAACCCATCTCTTGGATCTTCTTCAAATGGTGATGAT 1161  
Indels: 88  
26 ArgGlnGluLeuGlnMetLysLysSerLysLeuValThrMetLeuAspGluValGluGln 45  
QY 1162 CGATATAGTCATTGGGTAGATGAGATTCATACGGTTATATCAGGTTCCATGCTGCAACC 1221  
Indels: 88  
46 ArgTyrArgGlnTyrHisGlnMetGlnValValIleThrSerPheGluGlnAlaAla 65  
QY 1222 GAGTTAGAT---CCACAGTTACACACCGGTTTGCCTCCAAACCGCTTCTCTTTATAC 1278  
Indels: 88

Best Local Similarity: 29.56% Mismatches: 152  
Query Match: 13.49% Indels: 88  
DB: 12 Gaps: 17

US-09-423-575-1 (1-1886) x US-10-424-599-180416 (1-538)

QY 586 CTTTCATCCCTTGGATCAITTAAGACACATATGATATCC---TCAACAACATCGTGGGT 642  
Db 2 LeuAsnThrLeuHisLeuLeuLeuHisHisAspGlnArgHisLeuTrp---20  
QY 643 TTTGAAGCAAAATAGTGAGTTTCAGGCATTTTCAGGTGTAGTGTGTCAGAGTAA-----696  
Db 21 -----PheGluLeuThrProSerValHisProProSerSerThrThr 34  
QY 697 -----CMAATGATG-----TCTACATTCGGTGAAGAA 723  
Db 35 ThrValProGlnIleThrLysThrThrGlnArgAlaThrAsnSerAlaLeuGlyGln---53  
QY 724 GATTTCCCGCTTTCTAATTTCCGAATAAAGAACAAATGAGCTTTTCATTGAGCTTCATGATCA 783  
Db 54 -----ProCysLeuLeuSerLeuAlaLeuSerProLeuLeuAlaThrSerLeuLeuArg 71  
QY 784 GATGTTTCTGATGAATGCTCGGAGATAAGTCTT-----TGTGCGAGCTACAAGATTAGCC 837  
Db 72 GluPheGlySerAlaAlaTrpAlaArgHisProValTrpGlyValProValAlaArgLeuAla 91  
QY 838 TCAGACACAGCTTCITGCAGCAGCAAGACACATTTCTAATAACGTTGTACTCAAGTTTC 897  
Db 92 GlnGlnMetLeuSerGlnAspGlyTyrValSerSerAsnSerLysValAlaSerValTyr 111  
QY 898 TCTCAA-----CTTATATTTGGCTTCAAAATACCTTCACCTCTGTTCAAGAAATACTA 948  
Db 112 GlnGlnGlyHisPheLeuValLysAsnSerLysPheLeuValProAlaGlnAspLeuLeu 131  
QY 949 TCTCATTTCGCGCATCTCGCTCGATATTATCTCTCGAGAACCGAGTCAGAGCTGCT 1008  
Db 132 AsnGluPheCysSer-----LeuCysAlaLysGlnSerAspLeuGlyLysPro 147  
QY 1009 AGTTCAGCGCTTTACTTCACGTTTGGAGAATAATACTGAGTTCTTGTGATGATCTAAT 1068  
Db 148 ThrLysSerLeuLysLysGlnTrpGlu-----AspGlnGlu 159  
QY 1069 ACTCGAGCGGGTTTCGGATCTACATTTCAAAGAGACATAGAA-----1116  
Db 160 AsnAsnGlyValGlySerSerLysLysHisSerLeuThrSerLeuGluPheValGluLeu 179  
QY 1117 ---GCAAGAAACCACTCTCTGATCTCTTCAAATGGTGGATGATCGATATAGTCAT 1173  
Db 180 GlnLysArgLysThrLysLeuLeuSerMetLeuGluLeuValAspArgTyrLysHis 199  
QY 1174 TCGGTAGATGATTCATACGGTTATATACGGTTCATGCT---GCAACCGAGTTAGAT 1230  
Db 200 TyrArgAsnGlnMetLysSerValValSerSerPheGluAlaValAlaGlyAsnGlyAla 219  
QY 1231 CCACAGTTACACACCGGTTTGCCTCCAAACCGTTTCTTCTTATACAAAGACCTGAGA 1290  
Db 220 AlaThrValTyrSerAlaLeuAlaLeuLysAlaMetSerArgHisPheArgCysLeuLys 239  
QY 1291 GAGAGAAATCTCGAAGAAGATA-----ATC 1314  
Db 240 AspGlyIleLeuSerGlnIleGlnAlaThrArgLysAlaMetGlyGluLysAspProVal 259  
QY 1315 TCTATGGCATCTGATTGGAGAGGCGAAA-----GACAGACT 1353  
Db 260 AlaProGlyThrThr-----ArgGlyGluThrProArgLeuLysValIleAspGlnThr 277  
QY 1354 CAAGAAACCTCTATGTTCCACGACATGCTCTCTTCAGCGAGCTGAAACGAAAGAACCAT 1413  
Db 278 -----LeuArgGlnGlnArgAlaPheGlnGlnMetSerMetMetGluThr 292  
QY 1414 CAGATTTCGAGACCTCAACGAGGTTTCGCTGAGAAATCTGTTTCGTTCTACGGAATGG 1473  
Db 293 HisProNtrpArgProGlnArgGlyLeuProGluArgAlaValSerValLeuArgAlaTrp 312

1474 ATGTTCCAAAATCTCTTACCTTACCGGAAAGATTCGGAGAAACATCTTCTTACGTATA 1533  
Indels: 88  
Db 313 LeuPheGluHisPheLeuHisProTyrProSerAspValAspLysHisIleLeuAlaArg 332  
QY 1534 CGAAGTGGCTTGACAAAGTCAAGTATCAAACTGGTTTATAAATCGCGGGTTAGGCTA 1593  
Db 333 GlnThrGlyLeuSerArgGlyGlnValSerAsnTrpPheIleAsnAlaArgValArgLeu 352  
QY 1594 TGAAGCCGATGATACAGAGATGATGCGGAATATGAAACAGAGAGCAAGCTCAATAAC---1650  
Db 353 TrpLysProMetValGluGluMetTyrLeuGluGluValLysAspProGluAsnAlaIle 372  
QY 1651 -----AGTCACATTCACCAACCGGA-----1671  
Db 373 AlaSerSerGluGlyAlaThrAspGlnAspAsnAspIleAsnProAsnAsnValGlnTyr 392  
QY 1672 ---CCAACCTCTCGAATGCCAAAATCTGTATGATGAGCAAGCAATGCATAAATAAGAC 1728  
Db 393 ProProProLeuSerSerArgSerGluAspGlnLysProSerLeuValArgIleAsp 412  
QY 1729 AACAAATGTGTTTACCACACTTTGTGATAAT 1758  
Db 413 SerGluCysAlaSerSerIleIleAsnAsn 422

RESULT 11  
US-10-425-114-43080  
; Sequence 43080, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingtong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 43080  
; LENGTH: 481  
; TYPE: PRT  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: 700684286\_FLI.pep  
US-10-425-114-43080

Alignment Scores:  
Pred. No.: 2,99e-34 Length: 481  
Score: 440.50 Matches: 97  
Percent Similarity: 63.43% Conservative: 40  
Best Local Similarity: 44.91% Mismatches: 60  
Query Match: 13.41% Indels: 19  
DB: 12 Gaps: 5

US-09-423-575-1 (1-1886) x US-10-425-114-43080 (1-481)

QY 1057 GGTGATCTTAATAACTCGGAGCGGTTTCGA-----TCTACATTTCAA 1101  
Db 6 GlyIuAsnAsnAsnAsnAspGlyGlyLysGlnGlyValGluLeuSerThrAlaGln 25  
QY 1102 AGGAGACATTAGAAGCAAGAAACCCATCTCTTGGATCTTCTTCAAATGGTGATGAT 1161  
Db 26 ArgGlnGluLeuGlnMetLysLysSerLysLeuValThrMetLeuAspGluValGluGln 45  
QY 1162 CGATATAGTCATTGGGTAGATGAGATTCATACGGTTATATCAGGTTCCATGCTGCAACC 1221  
Db 46 ArgTyrArgGlnTyrHisGlnMetGlnValValIleThrSerPheGluGlnAlaAla 65  
QY 1222 GAGTTAGAT---CCACAGTTACACACCGGTTTGCCTCCAAACCGTTTCTTCTTTATAC 1278





Db 398 AlaTrrPheGluHisPheLeuHisProTyrProLysAspSerGluLysLeuMetLeu 417  
 QY 1528 GCTATACGAAGTGGTTCACAGAAGTCAAGTATCAAACTGTTTATAAATCGCGGGT 1587  
 Db 418 AlaArgGlnThrGlyLeuThrArgSerGlnIleSerAsnTrpPheIleAsnAlaArgVal 437  
 QY 1588 AGCTATGAAGCCGATGATAGAGAGATGTCGGAAATGAACAAGAGGAGCTCAAT 1647  
 Db 438 ArgLeuTrpLysProMetIleGluAspMetTyr-----LysGluGluIleGly 453  
 QY 1648 AACAGTCACATTCACCCACGACCACTTCGAATGCCAAATCT 1695  
 Db 454 GluAlaAspLeuAspSerAsnSerSerSerAspAsnValProArgSer 469  
 RESULT 13  
 US-10-425-114-48619  
 ; Sequence 48619, Application US/10425114  
 ; Publication No. US20040034898A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jingdong  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Screen, Steven E.  
 ; APPLICANT: Tabaska, Jack E.  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; TITLE REFERENCE: 38-21(5313)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,114  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 48619  
 ; LENGTH: 335  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: LIB3592-001-B7\_FLI.pbp  
 US-10-425-114-48619  
 Alignment Scores:  
 Pred. No.: 5,71e-34 Length: 335  
 Score: 437.00 Matches: 105  
 Percent Similarity: 54.08% Conservative: 54  
 Best Local Similarity: 35.71% Mismatches: 109  
 Query Match: 13.31% Indels: 26  
 DB: 12 Gaps: 5  
 US-09-423-575-1 (1-1886) x US-10-425-114-48619 (1-335)  
 QY 802 TCGGAGATAGTCTTTGTCAGCTACAGATAGCTCAGAGAGAGCTTCTTCGAGCAGC 861  
 Db 9 AlaGlnAlaGlyMetAlaAlaAlaPheLeuSerProAspGlnAlaSerAlaAlaAla 28  
 QY 862 AAAGACATTTCTAATAACGTTCTTACTCAAGGTTCTCTCAACTATATTTGGCTCAAAA 921  
 Db 29 SerArgSerAlaGlnSerIleTyrValGln-----AsnSerArg 41  
 QY 922 TACCTTCACTCTGTTCAGAATACTACTCTCATTTTCGCC-----GCA 963  
 Db 42 TyrLeuLysAlaAlaArgGluLeuLeuAspGluValValAsnValGlnAspAlaIleLys 61  
 QY 964 TACTCGCTCGATTATTCATCTCGAGGACCGAGTCAGAGCTGCTAGTTCAGCCCTTTACT 1023  
 Db 62 ArgLysGlyAspLysSerGlnGlnGlyLysAspSerGlyGlyGlyGlyGlyGlu 81  
 QY 1024 TCAGCTTTTGAGATATACTAGTCTTTTGATGGTGTAT---TCTAATAACTCGGAGCG 1080  
 Db 82 GlyLysAspAlaGluThrSerAspGluLysAlaGlyGluHisGluGlyAsnSerAla 101  
 QY 1081 GGTTCGGATCTACATTTCAAGGAGCATTAGAACCAAGAAACCCATCTCTTGGAT 1140  
 Db 102 ProGluLeuSerProSerGluArgGlnAspLeuGlnAsnLysValSerAlaLeuMetAla 121

QY 1141 CTCTTCAAATGGTGATGATCGATATAGTCATTGCGTAGATGAGATTATACCGTTATA 1200  
 Db 122 LeuLeuAspGlnValAspArgLysTyrArgHisGlnMetGlnMetValMet 141  
 QY 1201 TCAGCGTTCCATGCTGCAACCGAGTTAGAT---CCACAGTTACACACCGGTTTGCCTC 1257  
 Db 142 SerSerPheAspAlaValAlaGlyAlaGlyAlaAlaArgProTyrThrAlaLeuAlaLeu 161  
 QY 1258 CAAACCGTTTCTCTTCTTATACAAAGACCTGAGAGAGATCTGCAAGAGATAATCTCT 1317  
 Db 162 GlnThrIleSerArgHisPheArgSerLeuArgAspAlaValGlyAlaGlnValGlnSer 181  
 QY 1318 ATGGGATCTGTATTTGGAGAGAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1377  
 Db 182 LeuArgSerLeuGlyLysAspGlySerAlaGlnGlyGlyGlyLeuProArgLeu 201  
 QY 1378 CATTGCTTCTTCAGCAGCTGAAACGAAAGAAC----- 1410  
 Db 202 ArgTyrIleAspGlnGlnLeuArgGlnGlnArgAlaMetGlnGlnPheGlyMetMetGln 221  
 QY 1411 -----CATCAGATTGGAGACCTCAACGAGTTTCCTGAGAAATCTCTTCGGTTCTA 1464  
 Db 222 GlnProGlnHisAlaTrpArgProGlnArgGlyLeuProGluSerAlaValSerValLeu 241  
 QY 1465 CGGAATTTGGATGTTCCAAACTTCTTCCCTTACCCCTTACCCGAAAGATTCGGAGAAACATCT 1524  
 Db 242 ArgAlaTrpLeuPheLysPheLeuHisProTyrProLysAspSerGluLysLeuMet 261  
 QY 1525 CTAGCTATACGAAGTGGCTTGACAGAAGTCAAGTATCAAACTGGTTTATAAATCGCGG 1584  
 Db 262 LeuAlaArgGlnThrGlyLeuSerArgGlyGlnValSerAsnTrpPheIleAsnAlaArg 281  
 QY 1585 GTTAGGCTATGGAAGCCGATGATAGAGAGATGTCGCGAA 1626  
 Db 282 ValArgLeuTrpLysProMetIleGluGluMetTyrArgGlu 295  
 RESULT 14  
 US-10-437-963-143490  
 ; Sequence 143490, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(5322)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 143490  
 ; LENGTH: 647  
 ; TYPE: PRT  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MET4530\_44394C.1.pbp  
 US-10-437-963-143490  
 Alignment Scores:  
 Pred. No.: 2,65e-33 Length: 647  
 Score: 431.50 Matches: 149  
 Percent Similarity: 44.82% Conservative: 89  
 Best Local Similarity: 28.06% Mismatches: 218  
 Query Match: 13.14% Indels: 75  
 DB: 16 Gaps: 17  
 US-09-423-575-1 (1-1886) x US-10-437-963-143490 (1-647)



30	ArgAsnGluGlnSerTrpArgProGlnArgGlyLeuProGluLysValSerVal	49
1462	CTACGGAATTGGATGTTCCAAAATTCCTTCACCCTTACCGAAAGATTTCGGAGAACAT	1521
50	LeuArgAlaTrpMetPheGlnAsnPheLeuHisProTyrProLysAspAsnGluLysHis	69
1522	CTTCTAGCTATACGAAGTGGCTTGACAAGAAGTCAGGTATCAAACTGGTTTAAATGCG	1581
70	LeuLeuAlaIleGlnSerGlyLeuThrArgSerGlnValSerAsnTrpPheIleAsnAla	89
1582	CGSGTTTAGCTATGGAAGCGGATGATAGAAGAGATGTATCCGGAATATGAACAAGCG	1638
90	ArgValArgLeuTrpLysProLeuIleGluGluMetTyrSerGluValAsnLysLysAsn	109
1639	-----AAGCTCAATAAACAGTCAC	1656
110	ArgProIleGluLysSerGlyCysAspSerArgAsnTyrAlaSerIleGlyAsnGlnArg	129
1657	ATTCAACCCCAAC	1688
130	IleGlnMetAsn	133

Search completed: September 2, 2004, 12:39:23  
Job time : 356.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2004, 00:11:20 ; Search time 754 Seconds  
(without alignments)  
10626.129 Million cell updates/sec

Title: US-09-423-575-1  
Perfect score: 1886  
Sequence: 1 attagttataaatgttc.....ttaaaaaaaaaaaaaaaaaa 1886

Scoring table: IDENTITY NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 337863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq 29Jan04: \*  
1: geneseqn1980s: \*  
2: geneseqn1990s: \*  
3: geneseqn2000s: \*  
4: geneseqn2001as: \*  
5: geneseqn2001bs: \*  
6: geneseqn2002s: \*  
7: geneseqn2003as: \*  
8: geneseqn2003bs: \*  
9: geneseqn2003cs: \*  
10: geneseqn2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1882.8	99.8	1886	2	AAX01422	Aax01422 A. thaliana
2	1476	78.3	9359	6	ABK11039	Abk11039 pVDH636 v
3	1476	78.3	9359	6	ABK10687	Abk10687 Transform
4	128.4	6.8	323	3	AAC56287	Aac56287 Pinus rad
5	127.4	6.8	343	3	AAC56856	Aac56856 Pinus rad
6	126.2	6.7	2025	3	AAC39138	Aac39138 Arabidops
7	123.6	6.6	316	3	AAC57132	Aac57132 Pinus rad
8	123.6	6.6	462	3	AAC56228	Aac56228 Pinus rad
9	123.6	6.6	2043	6	ABZ13532	Abz13532 Arabidops
10	123.6	6.6	2043	8	ADB23149	Adb23149 Envirome
11	123.6	6.6	2385	9	ADD30303	Add30303 Plant yle
12	122.2	6.5	529	3	AAC55946	Aac55946 Eucalyptu
13	121.8	6.5	1599	6	ABZ14349	Abz14349 Arabidops
14	121.2	6.4	833	3	AAC34112	Aac34112 Arabidops
15	119.8	6.4	1449	3	AAC43406	Aac43406 Arabidops
16	119.8	6.4	2031	3	AAC36460	Aac36460 Arabidops
17	114.4	6.1	1884	6	ABZ13100	Abz13100 Arabidops
18	114.4	6.1	1983	4	AAD06494	Aad06494 Arabidops
19	114.4	6.1	1983	9	ADC46618	Adc46618 Thalecres
20	114.4	6.1	1983	9	ADD30786	Add30786 Plant yle
21	113.4	6.0	2545	4	AAD06479	Aad06479 Arabidops
22	113.4	6.0	2545	9	ADE37296	Ade37296 Plant yle
23	112.2	5.9	2450	7	AAL55527	Aal55527 qSH-1 gen

24	110.6	5.9	2442	7	AAL55526	Aal55526 qSH-1 gen
25	110	5.8	1836	6	ABZ12846	Abz12846 Arabidops
26	110	5.8	2405	4	AAD06449	Aad06449 Arabidops
27	110	5.8	2405	9	ADE37116	Ade37116 Plant yle
28	107	5.7	404	7	ABX62621	Abx62621 Arabidops
29	107	5.7	1575	4	AAD06493	Aad06493 Arabidops
30	107	5.7	1824	3	AAC36745	Aac36745 Arabidops
31	103.2	5.5	1905	3	AAC39451	Aac39451 Arabidops
32	100.6	5.3	2131	6	ABL56175	Ab156175 Hordeum v
33	100.6	5.3	2141	6	ABL56177	Ab156177 Hordeum v
34	96.8	5.1	381	3	AAC55971	Aac55971 Eucalyptu
35	96.8	5.1	381	3	AAC56750	Aac56750 Eucalyptu
36	94.4	5.0	380	3	AAC56524	Aac56524 Eucalyptu
37	94.4	5.0	380	3	AAC55980	Aac55980 Eucalyptu
38	92.4	4.9	626	5	AAH87740	Aah87740 Peppermin
39	89.8	4.8	706	6	ABK82114	Abk82114 DNA encod
40	84.8	4.5	260	3	AAC56220	Aac56220 Pinus rad
41	84.8	4.5	260	3	AAC57111	Aac57111 Pinus rad
42	79.8	4.2	366	3	AAC56519	Aac56519 Eucalyptu
43	79.2	4.2	474	3	AAC56677	Aac56677 Eucalyptu
44	79.2	4.2	474	3	AAC55926	Aac55926 Eucalyptu
45	79	4.2	157	3	AAC56263	Aac56263 Pinus rad

## ALIGNMENTS

RESULT 1  
AAX01422  
ID AAX01422 standard; cDNA; 1886 BP.  
XX  
AC AAX01422;  
DT 26-APR-1999 (first entry)  
XX  
DE A. thaliana ATH1 homeobox gene.  
XX  
KW ATH1; homeobox gene; plant gene construct; flowering modification;  
KW promoter; shade avoidance response; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN WO9851800-A1.  
XX  
PD 19-NOV-1998.  
XX  
PF 14-MAY-1998; 98WO-IB000821.  
XX  
PR 14-MAY-1997; 97GB-00009789.  
PR 30-DEC-1997; 97GB-00027458.  
XX  
PA (VHAV-) VAN DER HAVE BV D J.  
XX  
PI Smeekens S, Weisbeek P, Proveniers M;  
XX  
WP1; 1999-070121/06.  
XX  
PT Plant gene constructs containing ATH1 - used to modify flowering or the  
PT shade avoidance response in plants.  
XX  
PS Claim 5; Fig 1; 53pp; English.  
XX  
CC This sequence represents the Arabidopsis thaliana ATH1 homeobox gene. The  
CC invention relates to a plant gene construct comprising a complete or  
CC partial DNA sequence coding for an ATH1 gene product under the control of  
CC a promoter functional in plants. The construct can be used to modify  
CC (either promote or retard) flowering in plants. The construct can also be  
CC used in a method of producing a transgenic plant that lacks the shade  
CC avoidance response of a wild type plant

SQ Sequence 1886 BP; 599 A; 373 C; 344 G; 570 T; 0 U; 0 Other;  
Query Match 99.8%; Score 1882.8; DB 2; Length 1886;

		Best Local Similarity 99.9%; Pred. No. 0;				Matches 1884; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
QY	1	ATTAGTATATAAATGTTCTATTTGTTGATCTAGTCTCTGAATCTTTTGTAGTGGGAG	60						
Db	1	ATTAGTATATAAATGTTCTATTTGTTGATCTAGTCTCTGAATCTTTTGTAGTGGGAG	60						
QY	61	ATGATGAAGATTATGAATTTCTTCAATGAAATTTGTAAGAAAGAAAGACATAGAGAGCT	120						
Db	61	ATGATGAAGATTATGAATTTCTTCAATGAAATTTGTAAGAAAGAAAGACATAGAGAGCT	120						
QY	121	GCGAATGAAGTACACTGTTCTTTTCACGGAGAAAGAAAGATTAATTAAGCATTTCTTT	180						
Db	121	GCGAATGAAGTACACTGTTCTTTTCACGGAGAAAGAAAGATTAATTAAGCATTTCTTT	180						
QY	181	CTTCAGTTTAAACACACATTTTGGAAATTTTGTATGTAATAATTTCTTTTGGAGGTTGT	240						
Db	181	CTTCAGTTTAAACACACATTTTGGAAATTTTGTATGTAATAATTTCTTTTGGAGGTTGT	240						
QY	241	GTGTCGAAATCTTCCCAAGGTTCTATCAAGAGAGAGATTAAGTTTTCATAGAAAC	300						
Db	241	GTGTCGAAATCTTCCCAAGGTTCTATCAAGAGAGAGATTAAGTTTTCATAGAAAC	300						
QY	301	CCAATGGACAAAC	360						
Db	301	CCAATGGACAAAC	360						
QY	361	CAAAATCTCTCTCATGATTTTATACCTTCAAGAGAGAGATTCAACTTTCTTCAACA	420						
Db	361	CAAAATCTCTCTCATGATTTTATACCTTCAAGAGAGAGATTCAACTTTCTTCAACA	420						
QY	421	ATGCTTCCATGGAATACCATACAGATCAGATCTCTTACAAATGGGGGCTTTGATATTTT	480						
Db	421	ATGCTTCCATGGAATACCATACAGATCAGATCTCTTACAAATGGGGGCTTTGATATTTT	480						
QY	481	AATTCCTCTCTCACTAAACAAATACCTTCACTCTTCCACAGCTCTATCGATGTTCAAGAT	540						
Db	481	AATTCCTCTCTCACTAAACAAATACCTTCACTCTTCCACAGCTCTATCGATGTTCAAGAT	540						
QY	541	AACCGAATGTTGAGTTCAATGCTCTCTCTCTCATCTCTCTTCACTTCACTCTTGGAT	600						
Db	541	AACCGAATGTTGAGTTCAATGCTCTCTCTCTCATCTCTCTTCACTTCACTCTTGGAT	600						
QY	601	CATTTAAGACACTATGATGTTCTTCAACACACATGTCGGGTTTGGAGCAATAGTGAG	660						
Db	601	CATTTAAGACACTATGATGTTCTTCAACACACATGTCGGGTTTGGAGCAATAGTGAG	660						
QY	661	TTTCAGGCAATTTTCAGGTGTAGTTGGTCCAAAGTGAACCAATGATGTTCACTTCCGTTGAA	720						
Db	661	TTTCAGGCAATTTTCAGGTGTAGTTGGTCCAAAGTGAACCAATGATGTTCACTTCCGTTGAA	720						
QY	721	GAGATTTCCGTTTCTAATTTGATTAAGAAACATGAGCTTCTATGAGTCTTGCA	780						
Db	721	GAGATTTCCGTTTCTAATTTGATTAAGAAACATGAGCTTCTATGAGTCTTGCA	780						
QY	781	TCAGATGTTTCTGATCAATGCTCGGAGATAAGTCTTTGTCAGGCTACAAGATTAGCCTCA	840						
Db	781	TCAGATGTTTCTGATCAATGCTCGGAGATAAGTCTTTGTCAGGCTACAAGATTAGCCTCA	840						
QY	841	GAGCAAGCTTCTTGAGAGCAAAAGACATTTCTAATAACGTTTGTACTCAAGGTTTCTCT	900						
Db	841	GAGCAAGCTTCTTGAGAGCAAAAGACATTTCTAATAACGTTTGTACTCAAGGTTTCTCT	900						
QY	901	CAACTTATTTTGGCTCAAAATACCTTCACTCTGTTCAAGAAATACATCTCATTTTCCGC	960						
Db	901	CAACTTATTTTGGCTCAAAATACCTTCACTCTGTTCAAGAAATACATCTCATTTTCCGC	960						
QY	961	GCATCTCGCTGATTAATTCATCTCGAGAAACCGAGTCAGGAGTCGTAGTTCAGCCTTT	1020						
Db	961	GCATCTCGCTGATTAATTCATCTCGAGAAACCGAGTCAGGAGTCGTAGTTCAGCCTTT	1020						
QY	1021	ACTTCACGTTTGAAGATATACGATTTCTTGTATGTTGATTTCTAATACTCGGAGCG	1080						

Db	1021	ACTTCACGTTTTCAGAAATATAACTGATGTTCTTGTATGTTGATTTCTAATACTCGGAGCG	1080
QY	1081	GTTTCGGATCTACATTTCAAGAGAGAGCATTTAGAGCAAGAAAGAAACCCATCTCTTGGAT	1140
Db	1081	GTTTCGGATCTACATTTCAAGAGAGAGCATTTAGAGCAAGAAAGAAACCCATCTCTTGGAT	1140
QY	1141	CTTCTTCAAAATGCTGATGATCGATATAGTCAATGCTAGATGAGATTTCATCGGTTATA	1200
Db	1141	CTTCTTCAAAATGCTGATGATCGATATAGTCAATGCTAGATGAGATTTCATCGGTTATA	1200
QY	1201	TCAGCGTTCATCTGCAACCGAGTTAGATCCACATTCACACCCCGTTTCCCTCCAA	1260
Db	1201	TCAGCGTTCATCTGCAACCGAGTTAGATCCACATTCACACCCCGTTTCCCTCCAA	1260
QY	1261	ACCGTTTCTCTTATACAAAGAACCTGAGAGAGAGATCTGCAAGAAAGATTAATCTCTATG	1320
Db	1261	ACCGTTTCTCTTATACAAAGAACCTGAGAGAGAGATCTGCAAGAAAGATTAATCTCTATG	1320
QY	1321	GGATCTGTATTGAGAGAGGCAAGACAGACTCAAGAAACCTCTATGTTCCACAGCAT	1380
Db	1321	GGATCTGTATTGAGAGAGGCAAGACAGACTCAAGAAACCTCTATGTTCCACAGCAT	1380
QY	1381	TGCTTTCTTCAGCAGCTGAAACGAAAGAACCATTCAGATTTGGAGACCTCAACGAGGTTG	1440
Db	1381	TGCTTTCTTCAGCAGCTGAAACGAAAGAACCATTCAGATTTGGAGACCTCAACGAGGTTG	1440
QY	1441	CTTGAGAAATCTGTTTCGGTTCTACGGAATTTGATGTTCCAAAACCTTCTTCAACCTTAC	1500
Db	1441	CTTGAGAAATCTGTTTCGGTTCTACGGAATTTGATGTTCCAAAACCTTCTTCAACCTTAC	1500
QY	1501	CCGAAAGATTCGAGAAACATCTTCTAGCTATACGAGTGGCTTGACAGAAAGTCAGGTA	1560
Db	1501	CCGAAAGATTCGAGAAACATCTTCTAGCTATACGAGTGGCTTGACAGAAAGTCAGGTA	1560
QY	1561	TCAACTGTTTATAAATCGCGGGTTAGGCTATGGAAGCCGATGATAGAAGATGAT	1620
Db	1561	TCAACTGTTTATAAATCGCGGGTTAGGCTATGGAAGCCGATGATAGAAGATGAT	1620
QY	1621	GCGAAATGAACAGAGGAGCTCAATAACAGTCATTAACAGTCATTAACAGGACCACTCTT	1680
Db	1621	GCGAAATGAACAGAGGAGCTCAATAACAGTCATTAACAGTCATTAACAGGACCACTCTT	1680
QY	1681	CGAATCCCAAAATCTGTTATGATGAGCCAAAGCAATGCAATAAGACAACTTGTGT	1740
Db	1681	CGAATCCCAAAATCTGTTATGATGAGCCAAAGCAATGCAATAAGACAACTTGTGT	1740
QY	1741	TACCAACTTGTGATAAATTAGCAATTTGCTATCTATGATGCTCCCAAAACCTTAAACCATG	1800
Db	1741	TACCAACTTGTGATAAATTAGCAATTTGCTATCTATGATGCTCCCAAAACCTTAAACCATG	1800
QY	1801	TAGGACTATCATTAGTATGTTTATATGTTATATACCTCTTATCTTTGACTATTTTC	1860
Db	1801	TAGGACTATCATTAGTATGTTTATATGTTATATACCTCTTATCTTTGACTATTTTC	1860
QY	1861	ATTTTTAAAAAATAAAAAAAAAAAAAA	1886
Db	1861	ATTTTTAAAAAATAAAAAAAAAAAAAA	1886

RESULT 2  
 ABK11039  
 ID ABK11039 standard; DNA; 9359 BP.  
 XX  
 AC ABK11039;  
 XX  
 DT 05-JUN-2002 (first entry)  
 XX  
 DE pVDH636 vector containing Arabidopsis Ath1 gene.  
 XX  
 KW pVDH636 vector; circular; cyclic; grass; plant; herbicide resistance;  
 KW baseball; cricket; football; golf; rugby; soccer; tennis; lawn; park;  
 KW athletic field; animal feedstuff; grass flowering inhibitor; Ath1; gene;  
 KW ds.







5765 ATCTCTTCAATGTTGGATGATCGATATAGTCAATTCGTTAGATGAGATTCATACGGTTA 5824  
 1199 TATCAGCGTTCATGCTCAACCGAGTTAGATCCACAGTTACACACCGGTTTGCCTCC 1258  
 5825 TATCAGCGTTCATGCTCAACCGAGTTAGATCCACAGTTACACACCGGTTTGCCTCC 5884  
 1259 AAACCGTTTCTCTTATACAAAGAACCTTGAGAGAGAGATCTGCAAGAGATAAATCTCTA 1318  
 5885 AAACCGTTTCTCTTATACAAAGAACCTTGAGAGAGAGATCTGCAAGAGATAAATCTCTA 5944  
 1319 TGGGATCTGTTATGAGAGAGAGCAAGACAGACTCAAGAACCTTATGTTCCACAGC 1378  
 5945 TGGGATCTGTTATGAGAGAGAGCAAGACAGACTCAAGAACCTTATGTTCCACAGC 6004  
 1379 ATTGCCTTCTCAGCAGCTGAAACGAAAGAACCATCAGATTGGAGACTCAACGAGTT 1438  
 6005 ATTGCCTTCTCAGCAGCTGAAACGAAAGAACCATCAGATTGGAGACTCAACGAGTT 6064  
 1439 TGCCTGAGAAATCTGTTTCGGTTCACGGAATGGAGTTTCAAAATCTCTTCAACCTT 1498  
 6065 TGCCTGAGAAATCTGTTTCGGTTCACGGAATGGAGTTTCAAAATCTCTTCAACCTT 6124  
 1499 ACCCGAAGATTCGGAGAGAAATCTCTAGCTATACGAGTGGCTTGACAGAGTCAAG 1558  
 6125 ACCCGAAGATTCGGAGAGAAATCTCTAGCTATACGAGTGGCTTGACAGAGTCAAG 6184  
 1559 TATCAAACTGTTTATAATGCGCGGTTAGCTATGGAAGCGGATGATAGAAGATGT 1618  
 6185 TATCAAACTGTTTATAATGCGCGGTTAGCTATGGAAGCGGATGATAGAAGATGT 6244  
 1619 ATGCGGAATGACAGAGAGAGCTCAATAACAGTCAATTAACCCACGACCAACTC 1678  
 6245 ATGCGGAATGACAGAGAGAGCTCAATAACAGTCAATTAACCCACGACCAACTC 6304  
 1679 TTGGAATGCAAAATCTGTTATGATGAGCGCAAGCAATGATATAAAGACAACTTGTG 1738  
 6305 TTGGAATGCAAAATCTGTTATGATGAGCGCAAGCAATGATATAAAGACAACTTGTG 6364  
 1739 TTTACAACTTGTGATATAGGCAATGCTACTC 1774  
 6365 TTTACAACTTGTGATATAGGCAATGCTACTC 6400

RESULT 4  
 AAC56287  
 ID AAC56287 standard; DNA; 323 BP.  
 XX  
 AC AAC56287;  
 XX  
 XX 25-JAN-2001 (first entry)  
 XX  
 DE Pinus radiata transcription factor DNA sequence #87.  
 XX  
 KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;  
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;  
 KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;  
 KW type 2 Cys2His2; CCAAT box element; MYB; ss.  
 XX  
 OS Pinus radiata.  
 XX  
 XX WO200053724-A2.  
 XX  
 XX 11-MAR-1999; 99US-00266513.  
 XX  
 XX 18-AUG-1999; 99US-0149485P.  
 XX  
 XX 14-SEP-2000.  
 XX  
 XX 09-MAR-2000; 2000WO-US006112.  
 XX  
 XX 11-MAR-1999; 99US-00266513.  
 XX  
 XX 18-AUG-1999; 99US-0149485P.  
 XX  
 XX (GENE-) GENESIS RES & DEV CORP LTD.  
 XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Wood M, McGrath A, Shenk MA, Glenn M;  
 PI WPI; 2000-579369/54.  
 XX  
 DR New isolated polynucleotide encoding a plant transcription factor for  
 XX producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
 PT having modified gene expression or modified activity of a polypeptide.  
 PT  
 XX Claim 1; Page 153; 747pp; English.  
 PS  
 XX The present invention relates to novel plant transcription factors from  
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding  
 CC sequence for one such transcription factor. The transcription factor may  
 CC be used to produce a plant having modified gene expression such as a  
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or  
 CC mahogany species or to modify the activity of a polypeptide in a plant.  
 CC The transcription factors of the present invention are members from the  
 CC following families of regulatory proteins: bZIP, bZIP family of G-box  
 CC binding factors, basic helix-loop-helix zipper,  
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2  
 CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements and  
 CC MYB  
 XX  
 XX Sequence 323 BP; 99 A; 67 C; 80 G; 77 T; 0 U; 0 Other;  
 SQ  
 Query Match 6.8%; Score 128.4; DB 3; Length 323;  
 Best Local Similarity 67.7%; Pred. No. 9.5e-22;  
 Matches 180; Conservative 0; Mismatches 86; Indels 0; Gaps 0;  
 QY 1370 TCACACGAGCTTCCTCTTTCAGCAGCTGAAACGAAAGACCATCAGATTGGAGACCTC 1429  
 Db 15 TACGACAAACGCGGCTTTCACCATTTAGATTGATGGAGACGACCTTGGGACCGC 74  
 QY 1430 AACGAGGTTTTCCTGAGAAATCTGTTTCGGTCTTACGGAATTCGATTTCCAAACTTC 1489  
 Db 75 AGAGAGGACTTCCTGAAAGCTGTTCTGTTCTGTCGATGTTGTTGAGCATTTTC 134  
 QY 1490 TTCACCTTACCGGAAAGATTCGGAGAAACATCTTCTAGCTATACGAGTGGCTTGCAA 1549  
 Db 135 TGCACCGGTATCCAACTGATGTCAGATAAGCATATATTGGCTAAAGCAACTGGCCTTCAA 194  
 QY 1550 GAAGTCAGGTATCAAACTGTTTATAAATCGCGGGTTAGGCTATGGAAGCCGATGATAG 1609  
 Db 195 GAAGTCAGGTATCAAACTGTTTATAAATCGCGGGTTAGGCTATGGAAGCCGATGATAG 1609  
 QY 1610 AAGAGATGATGCGGAATGAAACAAG 1635  
 Db 255 AGGAGATGATGCGGAATGAAACAAG 280  
 RESULT 5  
 AAC56856  
 ID AAC56856 standard; DNA; 343 BP.  
 XX  
 AC AAC56856;  
 XX  
 XX 25-JAN-2001 (first entry)  
 DT  
 XX Pinus radiata transcription factor DNA sequence #302.  
 DE  
 XX  
 XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;  
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;  
 KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;  
 KW type 2 Cys2His2; CCAAT box element; MYB; ss.  
 XX  
 OS Pinus radiata.  
 XX  
 XX WO200053724-A2.  
 XX  
 XX 14-SEP-2000.  
 XX  
 XX 14-SEP-2000.

```

PF 09-MAR-2000; 2000WO-US006112.
XX
XX 11-MAR-1999; 99US-00266513.
PR 18-AUG-1999; 99US-0149485P.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Wood M, McGrath A, Shenk MA, Glenn M;
XX WPI; 2000-579369/54.
DR
XX
XX New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide.
XX
XX Claim 1; Page 530; 747pp; English.
PS
XX
XX The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper, LIM domain, AP2
CC homeotic/homeodomain/homeobox/HADS, homeodomain zipper, LIM domain, AP2
CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements and
CC MYB
XX
XX
SQ Sequence 343 BP; 106 A; 70 C; 86 G; 79 T; 0 U; 2 Other;

Query Match 6.8%; Score 127.4; DB 3; Length 343;
Best Local Similarity 67.3%; Pred. No. 1.7e-21;
Matches 179; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 1370 TCCACGAGATTGCGCTCTTTCAGCAGCTGAAACGAAAGAACCATCAGATTGGAGACCTC 1429
DB 35 TACGACACAGCGCCGCACTTCCACCACTTAGGATTGATGGAGCAGCACCCCTTGGCGACCGC 94

QY 1430 AACGAGGTTGGCTGAGAAAATCTGTTTCGGTTTACGGAATGATGTTCCAAACTTCC 1489
DB 95 AGAGAGGACTTCTCGAAGCGCTCTGTCTGTTCTTCGTGCGATGTTGTTGAGCATTTTC 154

QY 1490 TTCACCTTACCCGAAAGATTCCGAGAAACATCTTCTAGCTATACGAGTGGCTTGACAA 1549
DB 155 TCCACCGGTATCCACTGATGAGATAGCATATATTGGCTAAGCAAACTGGCCTTACAA 214

QY 1550 GAAGTCAGATATCAAACTGGTTTATAATGCGCGGTTAGCTATGGAAGCCGATGATAG 1609
DB 215 GAAGTCAGATATCAAACTGGTTTATAATGCGCGGTTAGCTATGGAAGCCGATGATG 274

QY 1610 AAGAGATGTATCGGAATGAACRAG 1635
DB 275 AGGAGATGTATCGGAAGAACTCAAG 300

RESULT 6
AAC39138
ID AAC39138 standard; DNA; 2025 BP.
XX
XX AAC39138;
AC
XX
XX 17-OCT-2000 (first entry)
DT
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 23512.
DE
XX Arabidopsis thaliana.
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX Arabidopsis thaliana.
OS

```

```

XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000BP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123348P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 08-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0133449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 08-MAY-1999; 99US-0132486P.
PR 08-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136382P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140153P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.

```

```
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143544P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144350P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148344P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 28-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 08-OCT-1999; 99US-0157753P.
PR 08-OCT-1999; 99US-0157865P.
PR 08-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 22-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161320P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 28-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 08-OCT-1999; 99US-0157753P.
PR 08-OCT-1999; 99US-0157865P.
PR 08-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 22-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161320P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
Query Match 6.7%; Score 126.2; DB 3; Length 2025;
Best Local Similarity 68.0%; Pred. No. 6.Se-21; Indels 0; Gaps 0;
Matches 176; Conservative 0; Mismatches 83;
QY 1368 GTTCCACCAGCATTCCTTCTTCAGCAGCTGAAACGAAAGAACCATCAGATTTGGAGACC 1427
Db 1196 GTTCAGACAGCAAGAGCTTTGCATCAGAGCTTGGATGGTTCCGCCCTTGGAGACC 1255
QY 1428 TCACGAGGTTGCTTCGAGAAATCTGTTTCGGTCTACGGAATGGATGTTCCAAACTT 1487
Db 1256 TCAAGAGGTTCTTCTGAAAACCTGTTTCTGTTCTCCGCTGCTTGGCTCTTCGACATTT 1315
QY 1488 CCTTACCCTTACCGGAAAGATTCCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGAC 1547
Db 1316 CCTTCATCCATATCCGAAAGAAATCTGAGAAAATCATGCTTCCGAAACAGACAGACTGTC 1375
QY 1548 AAGAAGTCAGGTATCAAACTGGTTTATAATGCGCGGTAGGCTATGGAAGCCGATGAT 1607
Db 1376 GAAGAACAGGTTCTTAATGGTTTCATAAACGCGCAGAGTTCTGTCTATGGAACCGATGAT 1435
QY 1608 AGAAGATCTATCGGAA 1626
Db 1436 CGAAGATGTATAAGAA 1454
RESULT 7
AAC57132
ID AAC57132 standard; DNA; 316 BP.
XX
AC AAC57132;
XX
```

```

DT 25-JAN-2001 (first entry)
DE Pinus radiata transcription factor DNA sequence #578.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Pinus radiata.
XX
XX WO200053724-A2.
XX
XX 14-SEP-2000.
XX
XX 09-MAR-2000; 2000WO-US006112.
XX
XX 11-MAR-1999; 99US-00266513.
XX
XX 18-AUG-1999; 99US-0149485P.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Wood M, McGrath A, Shenk MA, Glenn M;
XX
XX WPI; 2000-579369/54.
XX
XX New isolated polynucleotide encoding a plant transcription factor for
XX producing a plant e.g. a woody plant, preferably eucalyptus or pine,
XX having modified gene expression or modified activity of a polypeptide.
XX
XX Claim 1; Page 592-593; 747pp; English.
XX
XX The present invention relates to novel plant transcription factors from
XX Eucalyptus grandis or Pinus radiata. The present sequence is the coding
XX sequence for one such transcription factor. The transcription factor may
XX be used to produce a plant having modified gene expression such as a
XX woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
XX mahogany species or to modify the activity of a polypeptide in a plant.
XX The transcription factors of the present invention are members from the
XX following families of regulatory proteins: bZIP, bZIP family of G-box
XX binding factors, basic helix-loop-helix zipper,
XX homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
XX and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements and
XX MYB
XX
XX Sequence 316 BP; 100 A; 61 C; 84 G; 71 T; 0 U; 0 Other;
XX
Query Match 6.6%; Score 123.6; DB 3; Length 316;
Best Local Similarity 68.4%; Pred. No. 1.5e-20;
Matches 171; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 1386 TCTTCAGCAGCTGAAACGAAAGAACCATCAGATTGGAGACTCAACGAGGTTGCCTGA 1445
DB 7 TCTCCAACAATTAGCATGATTTCAGCAGCATGCTTGGAGGCCACAGAGAGACTTCCGA 66
QY 1446 GAAATCTGTTTCGGTCTTACGGAATGGATGTTCCAAAACCTTCTTACCCCTACCGAA 1505
DB 67 GCGATCTCTTTCTGCTTTCACGGGCTTGGCTATTTGAAACATTTCTTCGTTATCCAAA 126
QY 1506 AGATTCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAAGAGTCAGGTATCAAA 1565
DB 127 AGATCGACACACATATGCTCGGACACAGACTGGGCTTACCAAGATCAGGTCTCAA 186
QY 1566 CTGGTTTAAATGCGGGGTTAGCTATGGAACCGCATGATGAGAGATGTATGCGGA 1625
DB 187 TTGGTTTAAATGACCGTGTACGCTTGTGAGCCCTATGTTGAGAGATGTATGTTGA 246
QY 1626 AATGAACAAG 1635
DB 247 GGAACAAG 256

```

```

RESULT 8
AAC56228
ID AAC56228 standard; DNA; 462 BP.
XX
XX AAC56228;
XX
XX 25-JAN-2001 (first entry)
XX
DE Pinus radiata transcription factor DNA sequence #28.
XX
XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
XX poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
XX basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
XX homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;
XX type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
XX Pinus radiata.
XX
XX WO200053724-A2.
XX
XX 14-SEP-2000.
XX
XX 09-MAR-2000; 2000WO-US006112.
XX
XX 11-MAR-1999; 99US-00266513.
XX
XX 18-AUG-1999; 99US-0149485P.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Wood M, McGrath A, Shenk MA, Glenn M;
XX
XX WPI; 2000-579369/54.
XX
XX New isolated polynucleotide encoding a plant transcription factor for
XX producing a plant e.g. a woody plant, preferably eucalyptus or pine,
XX having modified gene expression or modified activity of a polypeptide.
XX
XX Claim 1; Page 138; 747pp; English.
XX
XX The present invention relates to novel plant transcription factors from
XX Eucalyptus grandis or Pinus radiata. The present sequence is the coding
XX sequence for one such transcription factor. The transcription factor may
XX be used to produce a plant having modified gene expression such as a
XX woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
XX mahogany species or to modify the activity of a polypeptide in a plant.
XX The transcription factors of the present invention are members from the
XX following families of regulatory proteins: bZIP, bZIP family of G-box
XX binding factors, basic helix-loop-helix zipper,
XX homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
XX and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements and
XX MYB
XX
XX Sequence 462 BP; 154 A; 85 C; 124 G; 99 T; 0 U; 0 Other;
XX
Query Match 6.6%; Score 123.6; DB 3; Length 462;
Best Local Similarity 68.4%; Pred. No. 1.7e-20;
Matches 171; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 1386 TCTTCAGCAGCTGAAACGAAAGAACCATCAGATTGGAGACTCAACGAGGTTGCCTGA 1445
DB 7 TCTCCAACAATTAGCATGATTTCAGCAGCATGCTTGGAGGCCACAGAGAGACTTCCGA 66
QY 1446 GAAATCTGTTTCGGTCTTACGGAATGGATGTTCCAAAACCTTCTTACCCCTACCGAA 1505
DB 67 GCGATCTCTTTCTGCTTTCACGGGCTTGGCTATTTGAAACATTTCTTCGTTATCCAAA 126
QY 1506 AGATTCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAAGAGTCAGGTATCAAA 1565
DB 127 AGATCGACACACATATGCTCGGACACAGACTGGGCTTACCAAGATCAGGTCTCAA 186
QY 1566 CTGGTTTAAATGCGGGGTTAGCTATGGAACCGCATGATGAGAGATGTATGCGGA 1625

```

```

Db      187 TTGGTTTATAATGACGTGACGCTCTGGAAGCTTGTGGAAGAGATGTATGGGA 246
QY      1626 AATGACCAAG 1635
Db      247 GGAACAAG 256

RESULT 9
ID ABZ13532
AC ABZ13532;
XX
DT 21-JAN-2003 (first entry)
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 1337.
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX WO200216655-A2.
XX 28-FEB-2002.
XX 24-AUG-2001; 2001WO-US026665.
XX 24-AUG-2000; 2000US-0227866P.
PR 26-JAN-2001; 2001US-0264647P.
PR 22-JUN-2001; 2001US-0300111P.
XX (SCRI ) SCRIPPS RES INST.
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI Harper JF, Krepes J, Wang X, Zhu T;
XX WPI; 2002-304127/34.
XX
PT Identifying a stress condition to which a plant cell has been exposed and
PT producing plants with increased tolerance to these abiotic stresses.
XX
PS Claim 144; SEQ ID NO 1337; 577pp + Sequence Listing; English.
XX
SS The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising: (a) contacting nucleic acid
CC representative of expressed polynucleotides in the plant cell with an
CC array or probes representative of the plant cell genome; and (b)
CC detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
CC in methods of the invention. Note: The sequence data for this patent is
CC not represented in the printed specification but is based on sequence
CC information supplied to Derwent by the European Patent Office
XX
SQ Sequence 2043 BP; 627 A; 507 C; 507 G; 402 T; 0 U; 0 Other;
Query Match 6.6%; Score 123.6; DB 6; Length 2043;
Best Local Similarity 69.4%; Pred. No. 2.9e-20;
Matches 168; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY 1410 CCATCAGATTGGAGACCTCAACGAGGTTTGGCTGAGAAATCTGTTCCGTTCTACGGAA 1469
Db 1146 CAATAATGCTTGGAGACCTCAACGTTGCTCTCCAGAACGAGCGCTCTCAGTTCTCCGTGC 1205
QY 1470 TTGGATGTTCCAAACTTCTTCCACCTTACCCGGAAGATTCCGGAACATCTTCTAGC 1529
Db 1206 TTGGCTCTTCGAACACTTCTTCTATCCATACCTTAGGATTCCGACACCATGCTAGC 1265
QY 1530 TATACGAAGTGGCTTGACAAGAGTCAGGTATCAAACTGGTTTATAAAATGCGCGGGTTAG 1589
Db 1326 GTTATGGAACCAATGTTGGAGGAGATGTATCATGAGGAAATGAAGGAGCGCAAGAA 1385

1266 TAAGCAACAGGACTCACTCGTAGCCAGGTGTCGAACTGGTTTATAAACGCGAGAGTTCCG 1325
1590 GCTATGGAAGCCGATGATAGAGAGATGATGCGGAAATGAACAAGAGAGCTCAATAA 1649
1326 GTTATGGAACCAATGCTGGAGGAGATGTACATGGAGGAAATGAAGGAGCGCAAGAA 1385
1650 CA 1651
1386 CA 1387

RESULT 10
ADB23149
ID ADB23149 standard; DNA; 2043 BP.
XX
AC ADB23149;
XX
DT 20-NOV-2003 (first entry)
DE Environmental stress-responsive promoter-related DNA, SEQ ID 153.
XX Environmental stress-responsive promoter; gene; ds.
XX Plant; environmental stress; promoter; gene; ds.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX WO2003044190-A1.
XX 30-MAY-2003.
XX 15-NOV-2002; 2002WO-JP011955.
XX 19-NOV-2001; 2001JP-00353038.
PR 29-JAN-2002; 2002JP-00020329.
XX (RIKE ) RIKEN KK.
XX Shinozaki K, Seki M, Fujita M;
XX WPI; 2003-468771/44.
DR P-PSDB; ADB23150.
XX Environmental stress-responsive promoters, useful in constructing and
XX breeding stress-resistant plants for applications in agriculture.
XX
PS Claim 10; Page 470-475; 520pp; Japanese.
XX
SS The present invention relates to novel environmental stress-responsive
CC promoters (ADB23163-ADB23252) from Arabidopsis thaliana. The promoters
CC are useful in constructing and breeding stress-resistant plants for
CC applications in agriculture and horticulture. The present sequence was
CC used to illustrate the invention.
XX
SQ Sequence 2043 BP; 627 A; 507 C; 507 G; 402 T; 0 U; 0 Other;
Query Match 6.6%; Score 123.6; DB 8; Length 2043;
Best Local Similarity 69.4%; Pred. No. 2.9e-20;
Matches 168; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY 1410 CCATCAGATTGGAGACCTCAACGAGGTTTGGCTGAGAAATCTGTTCCGTTCTACGGAA 1469
Db 1146 CAATAATGCTTGGAGACCTCAACGTTGCTCTCCAGAACGAGCGCTCTCAGTTCTCCGTGC 1205
QY 1470 TTGGATGTTCCAAACTTCTTCCACCTTACCCGGAAGATTCCGGAACATCTTCTAGC 1529
Db 1206 TTGGCTCTTCGAACACTTCTTCTATCCATACCTTAGGATTCCGACACCATGCTAGC 1265
QY 1530 TATACGAAGTGGCTTGACAAGAGTCAGGTATCAAACTGGTTTATAAAATGCGCGGGTTAG 1589
Db 1266 TAAGCAACAGGACTCACTCGTAGCCAGGTGTCGAACTGGTTTATAAACGCGAGAGTTCCG 1325
1590 GCTATGGAAGCCGATGATAGAGAGATGTATGCGGAAATGAACAAGAGAGCTCAATAA 1649
1326 GTTATGGAACCAATGTTGGAGGAGATGTATCATGAGGAAATGAAGGAGCGCAAGAA 1385
```

QY 1650 CA 1651  
Db 1386 CA 1387

RESULT 11  
ADD30303  
ID ADD30303 standard; cDNA; 2385 BP.  
XX  
AC ADD30303;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Plant yield-related polynucleotide clone G1589.  
XX  
KW ds; transcription factor; transgenic plant; growth rate; senescence;  
KW seed germination rate; plant vigor; seedling vigor.  
XX  
OS Arabidopsis thaliana.  
XX  
PN WO2003013227-A2.  
XX  
PD 20-FEB-2003.  
XX  
PF 09-AUG-2002; 2002WO-US025805.  
XX  
PR 09-AUG-2001; 2001US-0310847P.  
PR 19-NOV-2001; 2001US-0336049P.  
PR 11-DEC-2001; 2001US-0338692P.  
PR 14-JUN-2002; 2002US-00171468.  
XX  
PA (MEND-) MENDEL BIOTECHNOLOGY INC.  
XX  
PI Ratcliffe O, Riechmann JL, Adam LJ, Dubell AT, Heard JE;  
PI Pilgrim ML, Jiang C, Reuber TL, Creelman RA, Pineda O, Yu G;  
PI Brown PB;  
XX  
DR WPI; 2003-248221/24.  
DR P-PSDB; ADD30304.  
XX  
XX New plant transcription factor polynucleotides and polypeptides, useful  
PT in producing transgenic plants with commercially valuable properties,  
PT such as an alteration in a plant growth characteristic, e.g. growth rate  
PT or apomixis.  
XX  
PS Disclosure; SEQ ID NO 332; 454pp; English.  
XX  
CC The invention relates to a number of isolated Arabidopsis thaliana cDNA  
CC sequences and their encoded proteins which are especially transcription  
CC factor related cDNA's and proteins. The isolated or recombinant plant  
CC transcription factor polynucleotides and polypeptides are useful in  
CC producing transgenic plants with commercially valuable properties, i.e.  
CC modified or altered desirable traits as compared to a reference plant,  
CC such as an alteration in a plant growth characteristic, e.g. growth rate,  
CC germination rate of seeds, vigor of plants and seedlings, or leaf and  
CC flower senescence. Sequence information related to the polynucleotides  
CC and polypeptides can also be used in bioinformatic search methods. The  
CC transgenic plant is useful for growing a progeny plant from a parent  
CC plant. This sequence represents one of the cDNAs of the invention.  
XX  
SQ Sequence 2385 BP; 748 A; 578 C; 565 G; 494 T; 0 U; 0 Other;  
Query Match 6.6%; Score 123.6; DB 9; Length 2385;  
Best Local Similarity 69.4%; Pred. No. 3e-20; Indels 0; Gaps 0;  
Matches 168; Conservative 0; Mismatches 74;

QY 1410 CCATCAGATTGGAGACCTCAACAGAGTTTCCTCCAGAAATCTGTTTCGGTTCTACGGAA 1469  
Db 1324 CAATAATGCTTTGGAGACCTCAACAGTGGTCTCCAGACGAGCCGTCTCAGTTCTCCGTGC 1383  
QY 1470 TTGGATGTTCCAAACTTCTTACCCCTTACCCAAAGATTCGAGAAATCTTCTAGC 1529  
Matches 175; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Db 1384 TTGGCTTTCGACACTTTTTCATCATACCTCAAGGATTCGACAGCATGCTAGC 1443  
QY 1530 TATACGAAGTGGCTTGACAAGAGTCAAGTATCAAACTGGTTTATAAATGCGCGGTTAG 1589  
Db 1444 TAAGCAACAGGACTCACTCGTAGCCAGGTGTGAACTGTTTATAAACGCGAGATTGC 1503  
QY 1590 GCTATGGAAGCCGATGATAGAGAGATGTATGCGGAATGAACAAGAGGAAGTCAATAA 1649  
Db 1504 GTTATGGAACCAATGTTGAGGAGATGTACATGAGGAAATGACAGGACGACCAAGAA 1563  
QY 1650 CA 1651  
Db 1564 CA 1565

RESULT 12  
AAC55946  
ID AAC55946 standard; DNA; 529 BP.  
XX  
AC AAC55946;  
XX  
DT 25-JAN-2001 (first entry)  
XX  
DE Rucalypus grandis transcription factor DNA sequence #77.  
XX  
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;  
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;  
KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;  
KW type 2 Cys2His2; CCAAT box element; MYB; ss.  
XX  
OS Eucalyptus grandis.  
XX  
XX WO2000053724-A2.  
XX  
PD 14-SEP-2000.  
XX  
PF 09-MAR-2000; 2000WO-US006112.  
XX  
PR 11-MAR-1999; 99US-00266513.  
PR 18-AUG-1999; 99US-0149485P.  
XX  
XX (GENE-) GENESIS RES & DEV CORP LTD.  
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
XX  
XX Wood M, McGrath A, Shenk MA, Glenn M;  
PI WPI; 2000-579369/54.  
XX  
XX New isolated polynucleotide encoding a plant transcription factor for  
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
PT having modified gene expression or modified activity of a polypeptide.  
XX  
PS Claim 1; Page 64-65; 747pp; English.  
XX  
CC The present invention relates to novel plant transcription factors from  
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding  
CC sequence for one such transcription factor. The transcription factor may  
CC be used to produce a plant having modified gene expression such as a  
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or  
CC mahogany species or to modify the activity of a polypeptide in a plant.  
CC The transcription factors of the present invention are members from the  
CC following families of regulatory proteins: bZIP, bZIP family of G-box  
CC binding factors, basic helix-loop-helix zipper,  
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2  
CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements and  
CC MYB  
XX  
SQ Sequence 529 BP; 152 A; 127 C; 138 G; 112 T; 0 U; 0 Other;  
Query Match 6.5%; Score 122.2; DB 3; Length 529;  
Best Local Similarity 66.5%; Pred. No. 3.9e-20; Indels 0; Gaps 0;  
Matches 175; Conservative 0; Mismatches 88;

Query Match 6.5%; Score 121.8; DB 6; Length 1599;  
Best Local Similarity 71.9%; Pred. No. 7.3e-20;  
Matches 159; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 1368 GTTCACAGCATTCCTCTTTCAGCAGCTGAAACGAAAGAACCATCAGATTTCGAGACC 1427  
Db 52 GCTCAGGAGCAACAGACCCCTACAGAGCTTGGGATGATGCAAGCATGCATGAGGCC 111  
Qy 1428 TCAACGAGGTTGCTGTGAGAAATCTGTTTCGTTCTACGGAATTCGATGTTCCAAACTT 1487  
Db 112 GCAAGAGAGGACTCTCTGAGAGTCTGTTTCTATTCTTCGGGCTGGCTATTTCGACATT 171  
Qy 1488 CTTTCACCCCTTACCCGAAAGATTTCGAGAAACATCTTCTAGTATACGAAGTGGCTTGAC 1547  
Db 172 TCTTCATCTCTACCAAGGATCTGACAAATCTCTTTCGAGGAGACAGGCTTGAC 231  
Qy 1548 AAGAAGTCAAGTATCAAACTGTTTATAAATCGCGGTTAGGCTATGGAAGCCGATGAT 1607  
Db 232 AAGAAGTCAAGTCTGCAATGTTTCTCATCAATGCAAGAGTGGCTCTCTCGAAACCTATGGT 291  
Qy 1608 AGAAGAGATGATGCGGAATCA 1630  
Db 292 CGAAGAAATGTACAAAGAGAGA 314

RESULT 13  
ABZ14349  
ID ABZ14349 standard; DNA; 1599 BP.  
XX AC ABZ14349;  
XX DT 21-JAN-2003 (first entry)  
XX DE Arabidopsis thaliana stress regulated gene SEQ ID NO 2154.  
XX KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.  
XX OS Arabidopsis thaliana.  
XX PN WO200216655-A2.  
XX PD 28-FEB-2002.  
XX PF 24-AUG-2001; 2001WO-US026685.  
XX PR 24-AUG-2000; 2000US-0227866P.  
XX PR 26-JAN-2001; 2001US-0264647P.  
XX PR 22-JUN-2001; 2001US-0300111P.  
XX PA (SCRI ) SCRIPPS RES INST.  
XX PI (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX PI Harper JF, Kreps J, Wang X, Zhu T;  
XX WPI; 2002-304127/34.  
XX PT Identifying a stress condition to which a plant cell has been exposed and  
XX PT producing plants with increased tolerance to these abiotic stresses.  
XX PS Claim 144; SEQ ID NO 2154; 577pp + Sequence Listing; English.  
XX CC The invention relates to identifying a stress condition to which a plant  
XX CC cell has been exposed, comprising: (a) contacting nucleic acid  
XX CC representative of expressed polynucleotides in the plant cell with an  
XX CC array or probes representative of the plant cell genome; and (b)  
XX CC detecting a profile of expressed polynucleotides in the plant cell  
XX CC characteristic of a stress response. The method is useful in the  
XX CC production of transgenic plants, cells and seeds and in producing plants  
XX CC with increased tolerance to abiotic stress. The present sequence is that  
XX CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used  
XX CC in methods of the invention. Note: The sequence data for this patent is  
XX CC not represented in the printed specification but is based on sequence  
XX CC information supplied to Derwent by the European Patent Office  
XX CC Sequence 1599 BP; 481 A; 329 C; 405 G; 384 T; 0 U; 0 Other;

Query Match 6.5%; Score 121.8; DB 6; Length 1599;  
Best Local Similarity 71.9%; Pred. No. 7.3e-20;  
Matches 159; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 1405 AAGAACCATCAGATTTCGAGACCTCAACGAGGTTTCCTTGAGAAATCTGTTTCGGTTCTA 1464  
Db 928 ATGCAACCTCAGCTTCGAGACCTCAACGCGTCTACCTCGAAACCTCTGTTTGATCTC 987  
Qy 1465 CGGAATTCGATGTTCCAAACTTCCTTACCCCTTACCCGAAAGATTTCGGAAGAACATCTT 1524  
Db 988 CGTCTTGGCTCTTTCAGCATTTTCTTACCCCTTACCCAAAGGATTCTGATAAGATCATG 1047  
Qy 1525 CTAGCTATACGAAGTGGCTTGACAAAGATCAGGTATCAAACTGTTTATAAATGCGCGG 1584  
Db 1048 CTAGCAAGACAAACGGGCTTTCGAGCGGTCAGGTTTCAAACTGTTTCAAAATGACGA 1107  
Qy 1585 GTTAGGCTATGGAAGCCGATGATAGAGAGATCTATGCGGA 1625  
Db 1108 GTGCGTTTATGGAAGCAATGTTGGAGAGATATACAAGGA 1148

RESULT 14  
AAC34112  
ID AAC34112 standard; DNA; 833 BP.  
XX AC AAC34112;  
XX DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 5490.  
XX KW Hybridisation assay; genetic mapping; gene expression control;  
XX KW protein identification; signal transduction pathway; metabolic pathway;  
XX KW promoter; termination sequence; ss.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX PR 25-FEB-1999; 99US-0121825P.  
XX PR 05-MAR-1999; 99US-0123180P.  
XX PR 09-MAR-1999; 99US-0123548P.  
XX PR 23-MAR-1999; 99US-0125788P.  
XX PR 25-MAR-1999; 99US-0126284P.  
XX PR 29-MAR-1999; 99US-0126785P.  
XX PR 01-APR-1999; 99US-0127462P.  
XX PR 06-APR-1999; 99US-0128234P.  
XX PR 08-APR-1999; 99US-0128714P.  
XX PR 16-APR-1999; 99US-0129845P.  
XX PR 19-APR-1999; 99US-0130077P.  
XX PR 21-APR-1999; 99US-0130449P.  
XX PR 23-APR-1999; 99US-0130891P.  
XX PR 28-APR-1999; 99US-0131449P.  
XX PR 30-APR-1999; 99US-0132048P.  
XX PR 30-APR-1999; 99US-0132407P.  
XX PR 04-MAY-1999; 99US-0132484P.  
XX PR 05-MAY-1999; 99US-0132485P.  
XX PR 06-MAY-1999; 99US-0132486P.  
XX PR 06-MAY-1999; 99US-0132487P.  
XX PR 07-MAY-1999; 99US-0132863P.  
XX PR 11-MAY-1999; 99US-0134256P.  
XX PR 14-MAY-1999; 99US-0134218P.  
XX PR 14-MAY-1999; 99US-0134219P.  
XX PR 14-MAY-1999; 99US-0134221P.  
XX PR 18-MAY-1999; 99US-0134370P.  
XX PR 18-MAY-1999; 99US-0134768P.  
XX PR 19-MAY-1999; 99US-0134941P.  
XX PR 20-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 14-JUN-1999; 99US-0138847P.  
PR 16-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144083P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144321P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 22-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145222P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147360P.  
PR 09-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149168P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 26-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-015753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159285P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.



[illegible]

```
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145911P.
PR 28-JUL-1999; 99US-0145916P.
PR 02-AUG-1999; 99US-0145986P.
PR 02-AUG-1999; 99US-0145988P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151068P.
PR 27-AUG-1999; 99US-0151080P.
PR 27-AUG-1999; 99US-0151303P.
PR 30-AUG-1999; 99US-0151438P.
PR 31-AUG-1999; 99US-0151930P.
PR 01-SEP-1999; 99US-0152363P.
PR 07-SEP-1999; 99US-0153070P.
PR 10-SEP-1999; 99US-0153758P.
PR 13-SEP-1999; 99US-0154018P.
PR 15-SEP-1999; 99US-0154039P.
PR 16-SEP-1999; 99US-0154779P.
PR 20-SEP-1999; 99US-0155139P.
PR 22-SEP-1999; 99US-0155486P.
PR 23-SEP-1999; 99US-0157865P.
PR 24-SEP-1999; 99US-0158029P.
PR 28-SEP-1999; 99US-0158458P.
PR 28-SEP-1999; 99US-0158596P.
PR 29-SEP-1999; 99US-0157117P.
PR 04-OCT-1999; 99US-0157533P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.

PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 6.4%; Score 119.8; DB 3; Length 1449;
Best Local Similarity 87.3%; Pred. No. 2.2e-19; Mismatches 82; Indels 0; Gaps 0;
Matches 169; Conservative 0;

Qy 1375 CAGCATTGCCCTTCTTCAGCAGCTGAAACGAAACCATCAGATTTGGAGACCTCAACGA 1434
Db 811 CAACAAAGAGCGTTGCAGCGTTAGGTGTGATGCAACCTCACACTTGGCGGCTCAACGC 870
Qy 1435 GGTTCGCTGAGAAATCTGTTTCGGTTCTACGGAATGGATGTTCCAAACTTCCTTCAC 1494
Db 871 GGTTCACAGATTCTTCTGTTTGGTTCTCCGGTCTTGCTATTGAGCATTTCTCCAC 930
Qy 1495 CCTTACCCGAAAGATTTCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAGAACT 1554
Db 931 CCTTATCCAAAGATTTCAGACAAGATCATGCTCGCTAGACAAACGGGTTGAGCCGAGGC 990
Qy 1555 CAGGTATCAAACTGGTTTATAATGCGGGTTAGGCTATGGAAGCCGATGATAGAGAG 1614
Db 991 CAGGTATCGAACTGGTTTATAATGCGGCTGTGCTCTCTGGAACCCGATGTTGGAGGAG 1050
Qy 1615 ATGTATGCGGA 1625
Db 1051 ATGTACAAGGA 1061
```

Search completed: September 2, 2004, 06:19:12  
Job time : 769 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2004, 04:48:37 ; Search time 142 Seconds  
(without alignments)  
7370.686 Million cell updates/sec

Title: US-09-423-575-1

Perfect score: 1886

Sequences: 1 atttagttataaaattgtgc.....ttaaaaaaa

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/prodata/2/ina/5A-COMB.seq.\*

2: /cgn2\_6/prodata/2/ina/5B-COMB.seq.\*

3: /cgn2\_6/prodata/2/ina/6A-COMB.seq.\*

4: /cgn2\_6/prodata/2/ina/6B-COMB.seq.\*

5: /cgn2\_6/prodata/2/ina/PCTUS-COMB.seq.\*

6: /cgn2\_6/prodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	2.5	2305	4	US-09-976-594-1031
2	47.6	2.5	1248	4	US-09-489-847-101
3	45.2	2.4	1579	4	US-09-620-312D-927
4	44.4	2.4	780	4	US-09-134-000C-2237
5	44.4	2.4	832	4	US-09-621-978-2813
6	42	2.2	7218	1	US-08-232-463-14
7	41	2.2	2418	4	US-09-601-198-61
8	40.8	2.2	1770	4	US-09-976-594-195
9	40	2.1	998	3	US-09-122-400B-5
10	40	2.1	7218	1	US-08-232-463-14
11	39.8	2.1	1364	4	US-09-428-034-1
12	39.2	2.1	535	4	US-09-621-978-683
13	38.6	2.0	6113	4	US-10-204-708-13
14	38.4	2.0	2049	4	US-09-624-945-2
15	38.4	2.0	2110	3	US-09-419-459-1
16	38.4	2.0	2952	1	US-08-306-691B-54
17	38.2	2.0	6317	4	US-10-204-708-11
18	38.2	2.0	9098	4	US-09-358-082A-28
19	38	2.0	1398	4	US-08-956-171E-662
20	38	2.0	7424	4	US-09-596-002-8
21	37.8	2.0	816	4	US-09-134-000C-2630
22	37.6	2.0	6609	4	US-09-976-594-690
23	37.6	2.0	9636	4	US-08-323-170B-1
24	37.6	2.0	9636	4	US-08-954-441-1
25	37.4	2.0	1016	4	US-09-716-129-42
26	37.2	2.0	350	4	US-09-598-401C-23
27	37.2	2.0	832	4	US-09-621-978-2813

RESULT 1  
US-09-976-594-1031  
; Sequence 1031, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 143  
; SOFTWARE: PERL Program  
; SEQ ID NO 1031  
; LENGTH: 2905  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6673549 480885.2  
US-09-976-594-1031

Query Match 2.5%; Score 48; DB 4; Length 2905;  
Best Local Similarity 52.5%; Pred. No. 0.0056;  
Matches 105; Conservative 0; Mismatches 95; Indels 0; Gaps 0;  
QY 1471 TGGATGTTCCAAACTTCCTTCAACCTTACCCGAAAGATTTCGGAGAAACATCTTCTAGCT 1530  
Db 1130 TGGCTCTTCCAGCATCTCACACATCCGTACCTTCCGAGAGCAGAAACAGTTAGCG 1189  
QY 1531 ATACGAAGTGCTTGACAAGAAGTCAGGTATCAACTGGTTTATAATGCGCGGTAGG 1590  
Db 1190 CAAGACACAGACTTCAATCTCCAGTAACAACTGGTTTATTATGCCAAGAAGA 1249  
QY 1591 CTATGGAAGCCGATGATAGAGAGATGTATCGGAAATGAACAGAGAGCTCAATAAC 1650  
Db 1250 ATAGTACAGCCCATGATTGACCACTCAATCGAGCAGTGAAGCAGCAGCATATAGT 1309  
QY 1651 AGTCACATTCAACCCACGG 1670  
Db 1310 CCAGAGGGTCAGCCCATGGG 1329

RESULT 2  
US-09-489-847-101  
; Sequence 101, Application US/09489847  
; Patent No. 6476195  
; GENERAL INFORMATION:

```

; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: PZ031P1
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 101
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-489-847-101

Query Match          2.5%; Score 47.6; DB 4; Length 1248;
Best Local Similarity 55.4%; Pred. No. 0.0048;
Matches    92; Conservative   0; Mismatches 74; Indels   0; Gaps   0;

QY      163 AATAAGCATTATCTTCTTCCTAGTTTTTAACACACATTTTGAAATTTGTGATGTFAAAA 222
        |||||
Db       1067 AATACGTAGATTTTGGCTCTCTGTATGTGCACAACATTTTGTACATTATGTTATTGGA 1126

QY      223 TTCTCTTTGGAACGTTGTTGTTCTGAAATCTCCCAAGGTTCTATCAGAAGAGGAGG 282
        |||||
Db       1127 TTTTCTTTTCATACATTATATTCCTCTAAAACTCTCAAAAAAIAAAAAAAAAAAAAAAAA 1186

QY      283 ATAAAGTTTCATAGAAACCCCAATGCACACACACACACACACACAC 328
        |||||
Db       1187 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1232

RESULT 3
US-09-620-312D-927
; Sequence 927, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyao
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21

```

```

; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/03/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match 2.4%; Score 44.4; DB 4; Length 832;
Best Local Similarity 13.4%; Pred. No. 0.029;
Matches 50; Conservative 168; Mismatches 149; Indels 7; Gaps 1;

Qy 570 TCTCATCTCTCCACCTTCATCTTGGATCATTTAAGACACTATGATTCCTCAAA 629
Db 11 WYAKCWTKWSWSYMMYKTYWRRKKKAWKYKWTWYMYRYAMNGTYKK 70

Qy 630 CAACATGCGGGTTTGAAGCAATAGTACAGTTTCAGGCATTTTCAGGTAGTGTGCC 689
Db 71 KAVCRITKTKKKKGGYMMYWGWSYMMWTRTWGYAYRSMYWRKWKYAY 130

Qy 690 AAGTGAACCAATGATGTCTACATCTCGGTGAAGAAGATTTCCTCC-----GTTTCTAAATT 742
Db 131 RKTTCYSSKGTWKKWKKAWTTWKKTYWATRYMMWCMWTKRWRASWYCWGWKA 190

Qy 743 CGAATAAAGAAACATAGCTTTTATGAGCTCTTCATGATGATGATGATGATGATGCT 802
Db 191 RKWSTWRKRSYASARSAXRCYSCSWGMSWKYMWWRWGWATGAGMKAWRASCMMR 250

Qy 803 CGAGATAAGTCTTTGTGCGAGTCAAGATAGCTTCAGCAAGAGCTTCTTCACAGACA 862
Db 251 RYVAGSKTSYSKSMWCMWTRSHKYCYTKARWTGYCYKGGWGRGRWYASKYMKWRW 310

Qy 863 AAGACATTTCTAATACGTTGTACTCAAGGTTTCTCTCAACTTATATTTGGCTCAAAAT 922
Db 311 WWCWARMYRSTGTASMMWRWRYTMMKWKYAWARAAWMMWAWRACAAATA 370

Qy 923 ACCTTCACTCTGT 936
Db 371 TAATTATTATGCT 384

RESULT 6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible

; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/03/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match 2.2%; Score 42; DB 1; Length 7218;
Best Local Similarity 7.5%; Pred. No. 0.36;
Matches 33; Conservative 210; Mismatches 195; Indels 0; Gaps 0;

Qy 1287 GAGAGAGAGAAATCTGCAAGAAGATAATCTCTATGGGATCTGTATGGAGAGAGCAAGA 1346
Db 1431 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1372

Qy 1347 CAAGACTCAAGAACCTCTATGTTCCACGAGATTCCTTCACGAGCTGAACGAAA 1406
Db 1371 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1312

Qy 1407 GAACCATCAGATTGGAGACTCAAGAGCTTTCCTGAGAAATCTGTTCCGTTCTACG 1466
Db 1311 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1352

Qy 1467 GAATTGGATGTTCCAAAACCTTCCTTACCCTTACCCGAAAGATTCGGAGAGAACTCTCT 1526
Db 1251 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1192

Qy 1527 AGCTATAGAGTGGCTTGACAGAGCTCAGGTATCAAACTGTTTATAATCGCGGTT 1586
Db 1191 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1132

Qy 1587 TAGGCTATGGAAGCCGATGATAGAAGAGATGTATGCGGAAATGAACAAGAGAGCTCAA 1646
Db 1131 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1072

Qy 1647 TAACAGTCATCTCAACCCAAACGAGCAACTCTTCGAAATGCCAAATCTCTTATGATGAG 1706
Db 1071 RRRERATCGAAGCTCCCTCGAAGCTGAGCCAAAGCTGGGAATTAATCTCTGAGCGTATG 1012

Qy 1707 CCAAGCAATGCATAATA 1724
Db 1011 GCAACGAGAGAAAATA 994

RESULT 7
US-09-601-198-61
; Sequence 61, Application US/09601198
; Patent No. 5531583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.
; APPLICANT: Chen, Ellison Y.
```

; APPLICANT: Glass, Jennifer S.  
; APPLICANT: Glass, John I.  
; APPLICANT: Heiner, Cheryl R.  
; APPLICANT: Lefkowitz, Elliot  
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA  
; FILE REFERENCE: UAB-13452/22  
; CURRENT APPLICATION NUMBER: US/09/601,198  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR FILING DATE: 1998-01-30  
; NUMBER OF SEQ ID NOS: 181  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 61  
; LENGTH: 2418  
; TYPE: DNA  
; ORGANISM: Ureaplasma urealyticum  
US-09-601-198-61

Query Match 2.2%; Score 41; DB 4; Length 2418;  
Best Local Similarity 44.3%; Pred. No. 0.39; Mismatches 210; Indels 0; Gaps 0;  
Matches 167; Conservative 0;

QY 1 ATTAGTTATATAAAAGTTGCTATTTGTTGATCTAGTCTCTGAATCTTTTAGTGAGGCAG 60  
DB |||||  
142 ATTTTCATCAATAAAAGTTGCTATTAATGAACGATTAAACAAATCTTTATCAAAAGTTAG 201  
QY 61 ATGATCAAGATTATGAATTTCTTCTGAAATTTTGTAAAGAAACATAGAGAGCT 120  
DB |||||  
202 TAGGTTAAATTTTATTGTCATCTGTTTAAATATGCAATGCTGTTTAAATGCTAAT 261  
QY 121 GCGGAATGAAAGTACACTGCTCTTTCACGCGAAGAAAGATAAATAAGCATTTCTCTT 180  
DB |||||  
262 TCTGTTTAAAGGATTAAATTTGCTTTGAAATTAATAAATTTGCTCATGCTGTTTATC 321  
QY 181 CTTGAGTTTAAACACACATTTTGAATTTTGTGATGTAATAATCTCTTTGGAAGCTGT 240  
DB |||||  
322 AACAAATTTAAACAATGATAGTGTGTTGATCTAAATCTTTTGTGTTAAATCTAT 381  
QY 241 GTTGTCTGAAATCTTCCCAAGGTTCTATCAGAAAGAAAGATAAAGTTTCATAGAAAC 300  
DB |||||  
382 TTATTTCTTATTAAATCACTTTAGTAGGACAAATTAATCTTTATTTTATCATCAAT 441  
QY 301 CCAATGGACAAACAAACAAACAACTTTTGTGATATGTTGATATGTTGATATGTTGAT 360  
DB |||||  
442 AAAATGAGCAACAGCATCTTGTAGTTGTTTATTATTAAATATCATTAATGTTCAATAAT 501  
QY 361 CAAAATCTCTTCTCAT 377  
DB |||||  
502 TAAATTTCTAAATCAT 518

RESULT 8  
US-09-594-195  
; Sequence 195, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 195  
; LENGTH: 1770  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature

; OTHER INFORMATION: Incyte ID No. 6673549 995887.12  
; NAME/KEY: unsure  
; LOCATION: 846, 851  
; OTHER INFORMATION: a, t, c, g, or other  
US-09-976-594-195  
Query Match 2.2%; Score 40.8; DB 4; Length 1770;  
Best Local Similarity 55.7%; Pred. No. 0.38; Mismatches 62; Indels 0; Gaps 0;  
Matches 78; Conservative 0;

QY 1471 TGGATGTTCCAAATCTCTTACCCGAAAGATTCCGAGAAACATCTTCTAGCT 1530  
DB |||||  
1029 TGGTGTTCAGCACCTCTCCGACCCGCTACCCCTCGGAGGAGCAGAAACAGCTGGG 1088  
QY 1531 ATACGAAGTGGTTGACAAAGATCAGGTATCAAACTGTTTATAAATCGGCGGTAG 1590  
DB |||||  
1089 CAGGACACGGGCTCACCATCTCTGCAAGTCAACAACTGTTTATTAACCCCGGAGACGC 1148  
QY 1591 CTATGGAAGCCGATGATAGA 1610  
DB |||||  
1149 ATCGTCAACCTATGATCGA 1168

RESULT 9  
US-09-122-400B-5  
; Sequence 5, Application US/09122400B  
; Patent No. 6245974  
; GENERAL INFORMATION:  
; APPLICANT: Michalowski, Susan  
; APPLICANT: Spiker, Steven  
; TITLE OF INVENTION: MATRIX ATTACHMENT REGIONS  
; FILE REFERENCE: Michalowski and Spiker  
; CURRENT APPLICATION NUMBER: US/09/122,400B  
; CURRENT FILING DATE: 1998-07-24  
; PRIOR APPLICATION NUMBER: 60/066,118  
; PRIOR FILING DATE: 1997-08-06  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 998  
; TYPE: DNA  
; ORGANISM: Nicotiana tabacum  
US-09-122-400B-5  
Query Match 2.1%; Score 40; DB 3; Length 998;  
Best Local Similarity 47.0%; Pred. No. 0.48; Mismatches 124; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 147 ACGGAGAAAGAGATAAATAAGCATATCTCTTCTGAGTTGTTGTAATCTTCCCAAGGTT 266  
DB |||||  
420 ACTGAAAAAAGAGAGCTGAAAATCAATTTCTAAAGCAATTTTATTTGTAATAATCTGA 479  
QY 207 AATTTGATGTAATAATCTCTTGGAGCTGTTGTTGTAATCTTCCCAAGGTT 266  
DB |||||  
480 AAAAATCTAATAAATCTGAAAAATGAAATATTTTTTTCTAATTTTACAAAAA 539  
QY 267 TATCAGAAGAAGATATAAGTTTCATAGAACCAATGAGAACCAACAAACAAACAA 326  
DB |||||  
540 CTGCTTTAAAAAAGCTGAAAAATATTTCTAAACAAATATTTTGTAAAAATAAAAA 599  
QY 327 CACTTTTAGTCTCTGGATAATGTCATGACTACTAACCAAAATCTCTTCTCATGATTTAT 386  
DB |||||  
600 AATATTTTCTTCTTTTCTGATTTTATTTAGTTAAAAATATTTAAAGTTTTCCTGATTTT 659  
QY 387 ACCTCAAGAGAGATCAACTTC 410  
DB |||||  
660 AATTACTTTAGAAAAATTAATTTTCT 683

RESULT 10  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367



QY 1472 GGATGTTCCAAACTTCCTTCCCTTACCCCTTACCGAAGATTCCGAGAAACATCTTCTAGCTA 1531  
| | | | |  
Db 311 GGATGTTAAGCATCGGTTTAAGGCTACCTTTCAGAGAAGAGCAAAATGCTGTCAG 370  
| | | | |  
QY 1532 TACGAAGTGGCTTGACAGAAGTCAAGTATCAAACTGGTTTATAAATCGCG 1583  
| | | | |  
Db 371 AGAAGACCAATTTGCTTTGTTGCGGATTTCTTAAGTGGTTTATCAATGCTCG 422  
| | | | |

## RESULT 13

US-10-204-708-13/c  
; Sequence 13, Application US/10204708  
; Patent No. 6677731  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
; FILE REFERENCE: 5013.1012  
; CURRENT APPLICATION NUMBER: US/10/204,708  
; CURRENT FILING DATE: 2003-05-06  
; PRIOR APPLICATION NUMBER: PCT/EP01/03971  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 98  
; SEQ ID NO 13  
; LENGTH: 6113  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-13

Query Match 2.0%; Score 38.6; DB 4; Length 6113;  
Best Local Similarity 46.8%; Pred. No. 2.7;  
Matches 155; Conservative 0; Mismatches 174; Indels 2; Gaps 1;  
QY 189 TTTACACACATTTTGGAAATTTGATGTAAATCTCTTTGGAAACGTTGTGTCTG 248  
| | | | |  
Db 3637 TTTATCACTCAATTAATTTATTTTACCTAACCTTTTCTATACTTATTTAATATTTA 3578  
| | | | |  
QY 249 AAAT--CTTCCCAAGGTTCTATCAGAAGAAGAGATAAAGTTTCATAGAAACCCCAATG 306  
| | | | |  
Db 3577 AACTTATTTCCCTATAAATCTCTATCAACAATAATATATCTCACAAAAAAATAA 3518  
| | | | |  
QY 307 GACACACACAAACAAACACATTTTGTCTCTGGAATATGTCATGACTAACCAAAAT 366  
| | | | |  
Db 3517 AAAACCTCAAAAAACATAACATAATATTAATATTTCTTAACAAATTAACAAACAAAA 3458  
| | | | |  
QY 367 CCTCTTCTCATGATTTTATACCTTCAAGAGAAGATTCAACTTCTTCAACATGCTT 426  
| | | | |  
Db 3457 ATATTTCTTTTCCATAAATTTATCAATAAAATTAATAATATCAAACTATCAAACTAAAT 3398  
| | | | |  
QY 427 CCATGGAATACCATCAGATCAGATCCTCTACAAATGGGTGGCTTGTATATTTTCAATCT 486  
| | | | |  
Db 3397 TTTTAAATTAATAATATTAATTTACATACAATAACATTCACCCCTTTATATAATACTATTT 3338  
| | | | |  
QY 487 ATGCTGACTCAAAATACCTATCATCTTCTC 517  
| | | | |  
Db 3337 ATTATTTTAAATAATAATAATAAACCTC 3307  
| | | | |

## RESULT 14

US-09-624-945-2  
; Sequence 2, Application US/09624945

; Patent No. 6607915  
; GENERAL INFORMATION:  
; APPLICANT: Monia, Brett P.  
; APPLICANT: Wanciewicz, Edward  
; TITLE OF INVENTION: Antisense Modulation of E2A-Pbx1 Expression  
; FILE REFERENCE: ISPH-0477  
; CURRENT APPLICATION NUMBER: US/09/624,945  
; CURRENT FILING DATE: 2000-07-25  
; PRIOR APPLICATION NUMBER: 60/156,836  
; PRIOR FILING DATE: 1999-09-30  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2049  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-624-945-2

Query Match 2.0%; Score 38.4; DB 4; Length 2049;  
Best Local Similarity 58.9%; Pred. No. 1.8;  
Matches 66; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 1493 ACCCTTACCGAAGATTCCGAGAAACATCTTCTAGCTATACGAAGTGGCTTACAGAA 1552  
| | | | |  
Db 1133 ACCCTTACCGGAGTGGAGGAGCCAAAGAGGAGTTAGCCAAAGAGTGGCATCACAGTCT 1192  
| | | | |  
QY 1553 GTCAGGTATCAAACTGGTTTATAAATCGCGGGTTAGGCTATGGAAGCCGAT 1604  
| | | | |  
Db 1193 CCCAGGTATCAAACTGGTTTGAATAAAGCAATCCGTTACAAAGAAACAT 1244  
| | | | |

## RESULT 15

US-09-419-459-1/c  
; Sequence 1, Application US/09419459  
; Patent No. 6222028  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jin-Hao  
; APPLICANT: Cheng, Kuo-Joan  
; APPLICANT: Tsai, Cheng-Fang  
; APPLICANT: Chang, Chia-Chieh  
; TITLE OF INVENTION: CELLULOSE ENZYMES  
; FILE REFERENCE: 08919/037001  
; CURRENT APPLICATION NUMBER: US/09/419,459  
; CURRENT FILING DATE: 1999-10-15  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2110  
; TYPE: DNA  
; ORGANISM: Piromyces rhizinflata  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (3)...(1749)  
US-09-419-459-1

Query Match 2.0%; Score 38.4; DB 3; Length 2110;  
Best Local Similarity 49.0%; Pred. No. 1.9;  
Matches 102; Conservative 0; Mismatches 106; Indels 0; Gaps 0;  
QY 1679 TTCGAATGCCAAATCTGTTATGATGAGCCAAAGCAATGATATAAAGACAAATTTGTG 1738  
| | | | |  
Db 2062 TTTAAATTTTTTACACAAATTTTTATAATCTTATTAGCTATTATTATAAATACATATCCA 2003  
| | | | |  
QY 1739 TTTACCACTTTGTGATAAATAGGCAATTTGCTATGTTGCCAAACCTAAACCA 1798  
| | | | |  
Db 2002 TCTTTAAATTTCTTGATTTTAATTAATCTTTTTTTGCTTTTAAGTAAGAATAAAT 1943  
| | | | |  
QY 1799 TGTAGACTATCATTCAGTATGTTATAATGTPATAACAACCTCTTTTATCTTTGACTATT 1858  
| | | | |  
Db 1942 AATTAAATTTTATTTTCACTTATTATTATTTTATATTTTCTTTCTTATAAATTTT 1883  
| | | | |  
QY 1859 TCATTTTATTAATAAATAAATAAATAAATAA 1886  
| | | | |



Db 1882 TTTTATAAATAATAAGAAATAAAA 1855

Search completed: September 2, 2004, 08:25:14  
Job time : 144 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2004, 08:22:57 ; Search time 887 Seconds  
(without alignments)  
10568.913 Million cell updates/sec

Title: US-09-423-575-1  
Perfect score: 1886  
Sequence: 1 attagtataaaattgtgc.....ttaaaaaaaataaaaaaa 1886

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3267054 seqs, 2485319735 residues

Total number of hits satisfying chosen parameters: 6534108

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/1/pubpna/US05\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:  
14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:  
17: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:  
18: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:  
19: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171.6	9.1	422	13	US-10-424-599-50272
2	130.2	6.9	1555	13	US-10-425-114-13353
3	130.2	6.9	1995	17	US-10-437-963-7710
4	130.2	6.9	3033	13	US-10-424-599-130628
5	126.4	6.7	2799	17	US-10-437-963-31585
6	126.2	6.7	592	17	US-10-021-323-14220
7	125.2	6.6	1716	13	US-10-425-114-9882
8	124.4	6.6	1452	13	US-10-425-114-5047
9	124.4	6.6	1606	13	US-10-425-114-7804
10	124.2	6.6	1785	17	US-10-437-963-19673
11	123.6	6.6	2043	9	US-09-938-842A-1337
12	123.6	6.6	2043	11	US-09-938-842A-1337
13	123.6	6.6	2385	13	US-10-225-066A-335
14	123.6	6.6	2385	16	US-10-374-780A-2671

15	122.6	6.5	227	12	US-09-922-293-2000	Sequence 2000, Ap
16	121.8	6.5	1599	9	US-09-938-842A-2154	Sequence 2154, Ap
17	121.8	6.5	1599	11	US-09-938-842A-2154	Sequence 2154, Ap
18	121.4	6.4	1280	13	US-10-425-114-13349	Sequence 13349, A
19	121.4	6.4	1822	13	US-10-424-599-133558	Sequence 133558, A
20	119.2	6.3	3029	17	US-10-437-963-60354	Sequence 60354, A
21	119	6.3	1649	13	US-10-425-114-20370	Sequence 20370, A
22	115	6.1	3028	17	US-10-437-963-41007	Sequence 41007, A
23	115	6.1	3311	13	US-10-424-599-141176	Sequence 141176, A
24	114.8	6.1	1892	13	US-10-424-599-63874	Sequence 63874, A
25	114.4	6.1	1884	9	US-09-938-842A-905	Sequence 905, App
26	114.4	6.1	1884	11	US-09-938-842A-905	Sequence 905, App
27	114.4	6.1	1983	13	US-10-412-6998-637	Sequence 637, App
28	114.4	6.1	1983	13	US-10-225-066A-819	Sequence 819, App
29	114.4	6.1	1983	15	US-10-286-264-17	Sequence 17, Appl
30	114.4	6.1	1983	16	US-10-374-780A-2481	Sequence 2481, Ap
31	114.4	6.1	2008	13	US-10-424-599-21587	Sequence 21587, A
32	113.8	6.0	1637	13	US-10-425-114-13359	Sequence 13359, A
33	113.8	6.0	1645	13	US-10-424-599-65908	Sequence 65908, A
34	113.4	6.0	2545	13	US-10-412-6998-267	Sequence 267, App
35	113.4	6.0	2545	16	US-10-225-068-235	Sequence 235, App
36	113.4	6.0	2545	16	US-10-374-780A-2329	Sequence 2329, Ap
37	112.2	5.9	2517	17	US-10-437-963-3700	Sequence 3700, Ap
38	111.8	5.9	1179	13	US-10-425-114-7712	Sequence 7712, Ap
39	111.8	5.9	1546	13	US-10-425-114-29785	Sequence 29785, A
40	111.8	5.9	2736	13	US-10-424-599-37574	Sequence 37574, A
41	111.4	5.9	1237	17	US-10-437-963-50722	Sequence 50722, A
42	111	5.9	672	13	US-10-425-114-19291	Sequence 19291, A
43	111	5.9	1005	13	US-10-425-114-22835	Sequence 22835, A
44	111	5.9	1727	13	US-10-425-114-34958	Sequence 34958, A
45	111	5.9	2486	13	US-10-424-599-55280	Sequence 55280, A

## ALIGNMENTS

### RESULT 1

US-10-424-599-50272  
; Sequence 50272, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION: Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 50272  
; LENGTH: 422  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_16404C.1  
US-10-424-599-50272

Query Match	9.1%	Score 171.6;	DB 13;	Length 422;
Best Local Similarity	72.5%	Pred. No. 8.2e-34;		
Matches 222;	Conservative 0;	Mismatches 84;	Indels 0;	Gaps 0;
QY	1337	GAGCAAGACAAAGACTCAAGAAACCTCTATGTTCCACACAGATTGCTTTTCAGCAGC	1396	
DB	25	GAGAAAGATAGAACCTTTGAAACATCGTTTATCAGAAACAATGGCTTTTCAACAGC	84	
QY	1397	TGAACGAAGAACCATCAGATTGGAGACCTCAAGAGGTTTGGCTTCCAGAAATCTGTTT	1456	
DB	85	TGAGAGGAACGACGACAAATCTTGGAGGCTCAGAGGGGCTGCTGAGAAATCTGTCT	144	
QY	1457	CGGTTCTACGGAATTGATGTTCCAAACCTTCTTACCTTACCCGAAAGATTCGGAGA	1516	

Db 145 CGGTTCTACGTGATGTTCCAAACTTCTTCCACCGTACCCAAAGACAAACGAAA 204  
Qy 1517 AACATCTTCTAGCTATACGAGTGGCTTGACAAGAGTCAGGTATCAAACTGGTTTATAA 1576  
Db 205 AGCATTTGCTACCCATTCAAGTGGATTGACGAGAGCCAGGTGTCACACTGGTTTATAA 264  
Qy 1577 ATGCGGGGTTAGGCTATGGAAGCCGATGATAGAGAGATGTATCGGAAATGAACAGA 1636  
Db 265 ATGCGAGAGTTTCGCTCTGGAAACCATTTGATAGAGGAATGTACTCGGAGGTAAACAAA 324  
Qy 1637 GGAAGC 1642  
Db 325 AGAATC 330

## RESULT 2

US-10-425-114-13353  
; Sequence 13353, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 13353  
; LENGTH: 1555  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: JC-GMST02400042B10\_FLI  
US-10-425-114-13353

Query Match 6.9%; Score 130.2; DB 13; Length 1555;  
Best Local Similarity 67.5%; Pred. No. 1e-22;  
Matches 183; Conservative 0; Mismatches 88; Indels 0; Gaps 0;  
Qy 1370 TCACACAGCATTCCTTCTCAGCAGCTCAACGAAAGAACCATCAGATTTGGAGCCTC 1429  
Db 411 TCAGGCAACAAAGGCGACTTCAGCAGCTAGGAATGATTCACCCCAATGCATGGAGGCCCC 470  
Qy 1430 AACGAGGTTTGCTGAGAAATCTGTTTCGGTTCTACGGAATGGATGTTCCAAAACCTCC 1489  
Db 471 AAAGAGGCTTGCTGACGAGCTGTTTCCATTCTTCGGGCTTGGCTTTTGAGCATTTCC 530  
Qy 1490 TTCACCTTTACCGAAAGATTCGGAGAACATCTTCTAGCTATACGAGTGGCTTGACAA 1549  
Db 531 TTCACCTTTATCCAAAGGACTCCGATAAAGTATGCTTGTAAACAAACTGGACTTGCTA 590  
Qy 1550 GAAGTCAGTATCAAACTGGTTTAAATGCGCGGTTAGGCTATGGAAGCCGATGATAG 1609  
Db 591 GGAGCCAGGTGCAAACTGGTTTATCAATGCCGAGTTTCGCTTTGGAGGCAATGGTTG 650  
Qy 1610 AAGAGATGATCGGAAATGAACAGAGGAA 1640  
Db 651 AAGAAATGTACTTGAAGAAATCAAGGAGCA 681

## RESULT 3

US-10-437-963-7710/c  
; Sequence 7710, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5322)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204956  
; SEQ ID NO 7710  
; LENGTH: 1995  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_1427C.1  
US-10-437-963-7710

Query Match 6.9%; Score 130.2; DB 17; Length 1995;  
Best Local Similarity 65.1%; Pred. No. 1.2e-22;  
Matches 192; Conservative 0; Mismatches 103; Indels 0; Gaps 0;  
Qy 1342 AAGACAAGACTCAAGAAACCTCTATGTTCCACCAGCATTCCTCTTCAGCAGCTGAAA 1401  
Db 510 AAGAGCGGAGCTGGAGTGTGCTTCATCCAGAACCATGGCGATGCGAGCTCCGG 451  
Qy 1402 CGAAGAACCATCAGATTTGGAGACCTCAACGAGTTTCCTCGAGAAATCTCTTTGGTT 1461  
Db 450 CGAGGTGACGAGCATGCTGTGGCGCGCAGCGCGCTTCCCGAGAAAGTCCCGCGTG 391  
Qy 1462 CTACGGAATGGATGTTCCAAAACCTCTTCAACCTTACCCCGAAGATTCGAGAAACAT 1521  
Db 390 CTCAGGCTGGATGTTTCAGAACTTCTCCCGCGATTCCAAAGACAGCCGAGAGGAC 331  
Qy 1522 CTTCTAGCTATACGAAGTGGCTTGACAAGAAGTCAGGTATCAAACTGGTTTATAAATCG 1581  
Db 330 ATGCTGCGCGGAGAGCGCTGAGCAGGAGCCAGGTCTCAACTGGTTTCAAAACGCC 271  
Qy 1582 CGGTTAGGCTATGGAAGCCGATGATAGAGAGATGTATGCGGAATGACAGA 1636  
Db 270 CGCGTCCGACTGTGGAAGCCGATGATCGAGGACATGTACGAGGAGCTCAAGAAGA 216

## RESULT 4

US-10-424-599-130628  
; Sequence 130628, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5322)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285584  
; SEQ ID NO 130628  
; LENGTH: 3033  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_98966C.1  
US-10-424-599-130628

Query Match 6.9%; Score 130.2; DB 13; Length 3033;  
Best Local Similarity 67.5%; Pred. No. 1.6e-22;  
Matches 183; Conservative 0; Mismatches 88; Indels 0; Gaps 0;  
Qy 1370 TCCACCAGCATTCCTTCTCAGCAGCTGAAACGAAAGAACCATCAGATTTTGAGACCTC 1429

Db 1374 TCAGGCAACAAAGGCACTTCAGCAGCTAGGATGATTCACCCCAATGCATGGAGGCC 1433  
QY 1430 ACGAGGTTGGCTGAGAAATCTGTTTCGGTCTACGGAATGGAATGTTCCAAAACCTCC 1489  
Db 1434 AAAGAGGCTTGCTGAAAGAGCTGTTCCATCTTCGGGCTTGGCTTTTGAGCATTTCC 1493  
QY 1490 TTCACCTTACCCGAAAGAGTTCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAA 1549  
Db 1494 TTCACCTTATCCAAAGGACTCGATAAAGTTATGCTTCTTAAACAACTGGAATGTCTA 1553  
QY 1550 GAAGTCAGGTATCAAACTGGTTTAAATGCGCGGTTAGGCTATGGAAGCCGATGATAG 1609  
Db 1554 GGAGCCAGGTGTCAAACTGGTTTATCAATGCCGAGTTCGGCTTTCGGAAGCCAAATGGTTG 1613  
QY 1610 AAGAGATGTATCGGAAATGAACAGAGGAA 1640  
Db 1614 AAGAAATGACTTGGAGAAATCAAGGAGCA 1644

RESULT 5  
US-10-437-963-31585  
; Sequence 31585, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: Deikman, Jill  
; APPLICANT: Feng, Paul C.C.  
; APPLICANT: Fincher, Karen L.  
; APPLICANT: Ziegler, Todd E.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(52274)B  
; CURRENT APPLICATION NUMBER: US/10/021,323  
; CURRENT FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: US 60/255, 619  
; PRIOR FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 17880  
; SEQ ID NO 14220  
; LENGTH: 592  
; TYPE: DNA  
; ORGANISM: Gossypium hirsutum  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(592)  
; OTHER INFORMATION: unsure at all n locations  
; OTHER INFORMATION: Clone ID: LIB9829-031-Q6-K6-E4  
US-10-021-323-14220

Query Match 6.7%; Score 126.4; DB 17; Length 2799;  
Best Local Similarity 68.4%; Pred. No. 1.5e-21;  
Matches 175; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1375 CAGCATGCTCTTCAGCAGCTGAAACGAAAGACCAATCAGATTTGGAGACTTCAACGA 1434  
Db 1590 CAACAGCGGCTTTCAGCAGTATGTTTGTACAGCAAAATGCTTGGAGGCCACAGAGG 1649  
QY 1435 GGTTCCTCGAGAAATCTGTTTCGGTCTACGGAATGGAATGTTCCAAAACCTTCTTCAC 1494  
Db 1650 GGAATCGCGAAACTCAGTTCAATCTCTGCTTGGCTTTTGAACACTTCTTCAC 1709  
QY 1495 CTTACCCGAAAGATTCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAAGT 1554  
Db 1710 CCGTATCCAAAGATTCAGAAAGCTGCTAGCAGACAACTGGCTTAAACAAAGT 1769  
QY 1555 CAGGTATCAACTGGTTTAAATGCGCGGTTAGGCTATGGAAGCCGATGATAGAGAG 1614  
Db 1770 CAGATTTCAATTTGGTTTATAAATGCCCGTGTCCGCTGTGGAACCCGATGATCAAGAC 1829  
QY 1615 ATGTATCGGAAATGA 1630  
Db 1830 ATGTATAAGAGAGA 1845

RESULT 6  
US-10-021-323-14220

; Sequence 14220, Application US/10021323  
; Publication No. US20040123340A1  
; GENERAL INFORMATION:  
; APPLICANT: Deikman, Jill  
; APPLICANT: Feng, Paul C.C.  
; APPLICANT: Fincher, Karen L.  
; APPLICANT: Ziegler, Todd E.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(52274)B  
; CURRENT APPLICATION NUMBER: US/10/021,323  
; CURRENT FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: US 60/255, 619  
; PRIOR FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 17880  
; SEQ ID NO 14220  
; LENGTH: 592  
; TYPE: DNA  
; ORGANISM: Gossypium hirsutum  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(592)  
; OTHER INFORMATION: unsure at all n locations  
; OTHER INFORMATION: Clone ID: LIB9829-031-Q6-K6-E4  
US-10-021-323-14220

Query Match 6.7%; Score 126.2; DB 17; Length 592;  
Best Local Similarity 68.7%; Pred. No. 6e-22;  
Matches 173; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 1375 CAGCATGCTCTTCAGCAGCTGAAACGAAAGACCAATCAGATTTGGAGACTTCAACGA 1434  
Db 27 CAACANAGAGCTCTTCAGCAGTTCGATGATGATCAACATGATGATGAGAGCCCTCAGAGA 86  
QY 1435 GGTTCCTCGAGAAATCTGTTTCGGTCTACGGAATGGAATGTTCCAAAACCTTCTTCAC 1494  
Db 87 GGGCTGCCGAAAGCTCTGTTTCAATCTTCGCTTGGCTATTTGAGCATTTTCTTCAT 146  
QY 1495 CTTACCCGAAAGATTCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAAGT 1554  
Db 147 CCTTACCCGAGGATTCGATAGATCATGCTAGCAAGACAGACAGGCTTAACTAGAAGT 206  
QY 1555 CAGGTATCAACTGGTTTATAAATGCCCGGTTAGGCTATGGAAGCCGATGATAGAGAG 1614  
Db 207 CAGTTCGAATTTGTTTATAAATGCTCGGCTGCTTTTGGAGAGCCCATGGTGAAGAG 266  
QY 1615 ATGTATCGGAA 1626  
Db 267 ATGTACAGGA 278

RESULT 7  
US-10-425-114-9882  
; Sequence 9882, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 9882  
; LENGTH: 1716  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:

; OTHER INFORMATION: Clone ID: 700875318\_FLI  
US-10-425-114-9882

Query Match	6.6%	Score 125.2	DB 13	Length 1716
Best Local Similarity	66.9%	Pred. No. 2.2e-21		
Matches 178	Conservative 0	Mismatches 88	Indels 0	Gaps 0
QY	1370	TCACACGAGTATGCCCTCTTCAGCAGCTGAAACGAAAGAACCATCAGATTTGGAGACCTC	1429	
Db	213	TTTCGGCAACAAGGGCACTTCAGCACTAGGAATGATTCAGGCCAATGCATGGAGGCCCC	272	
QY	1430	AACGAGGTTTGGCTGAGAAATCTCTTTTCGGTTCCTAAGGAATTGGATGTTCCAAAACITCC	1489	
Db	273	AGAGAGGTTTGCCTGGAACGAGCTCTTTCCATCTTCGGGCTTGGCTTTTTCGACATTTCC	332	
QY	1490	TTACACCCCTTATCCCGAAAGATTTCGGAGAAACATCTTCTAGCTATACGAATGGCTTTGCAA	1549	
Db	333	TTACCCCATATCCAAAGSACTCCGATAAAGTTATGCTTTAAACAACATGGACITTTCTA	392	
QY	1550	GAACTCAGGTATCAAACTGGTTTTAAATCGCGGGTTAGGCTATCGAGCCGATGATAG	1609	
Db	393	GGAGCCAGGTGTCAAACTGGTTTTCAATCCCGAGTTCGGCTTTTGGAAAGCCAATGGTTG	452	
QY	1610	AAGAGATGTATCGGAAATGAACAAG	1635	
Db	453	AGAAGATGTACTTGGAGAAATCAAG	478	

```

RESULT 8
US-10-425-114-5047
; Sequence 5047, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313) B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 5047
; LENGTH: 1452
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700441126_FLI
US-10-425-114-5047

```

## RESULT 9

```

US-10-425-114-7804
; Sequence 7804, Application US/10425114
; Publication NO. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Jiu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53)313B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 7804
; LENGTH: 1606
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700694286_FLI
US-10-425-114-7804

```

RESULT 10

US-10-437-963-19673/C  
; Sequence 19673, Application US/10437963  
; Publication NO. US2004012343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 19673  
; LENGTH: 1785  
; TYPE: DNA

; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_2510C.1  
US-10-437-963-19673

Query Match 6.6%; Score 124.2; DB 17; Length 1785;  
Best Local Similarity 65.1%; Pred. No. 4.1e-21;  
Matches 183; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 1357 GAAACCTCTATGTTCCACAGAGTTGCCCTTCTTCAGCAGCTGAACGAAAGAACCATGAG 1416  
DB 399 GAGTCGGCGTTATCCAGAGAGCACTGGCGGGTGCAGCAGCTCGGCGCGGAGCAGCAG 340  
QY 1417 ATTGTGAGACCTCAACGAGGTTTGCTGAGAAATCTGTTTCGGTTCTACGGAATTGGATG 1476  
DB 339 TGTGCGCGCGAGCGCGCTCCGAGAGAGTCCGTCGCGTCTCAAGSCCTGGATG 280  
QY 1477 TTCCAAAATCTCTTCAACCTTACCCGAAAGATTGCGAAGAACATCTTCTAGCTATACGA 1536  
DB 279 TTGAGAATCTCTCCGCGCGTATCCGAAAGACAGCGAGAGGAGATGTCGCGAGG 220  
QY 1537 AGTGGCTTGACAGAGTTCAGGTATCAAACTGTTTATAAATGCGCGGTTAGGCTATGG 1596  
DB 219 AGCGGCTCAGCAGGAGCAGGTGTCGAATTGTTCTAATACGTCGAGTTCGGCTATGG 160  
QY 1597 AAGCCGATGATGAAGAGATGTATGCGGAAATGAACAAG 1637  
DB 159 AAGCCAATGATCAGGAGATGTCGAGGAGCTGAAGAGGAG 119

RESULT 11  
US-09-938-842A-1337  
; Sequence 1337, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: SCRIPT1300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 1337  
; LENGTH: 2043  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-1337

Query Match 6.6%; Score 123.6; DB 9; Length 2043;  
Best Local Similarity 69.4%; Pred. No. 6.3e-21;  
Matches 168; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1410 CCATCAGATTGGAGACCTCAACGAGGTTTGCTGAGAAATCTGTTTCGGTTCTACGGAA 1469  
DB 1146 CAATAATGCTTGGAGACCTCAACGAGGTTTGCTGAGAAATCTGTTTCGGTTCTACGGAA 1205  
QY 1470 TTGGATGTTCCAAAATCTCTTACCCCTTACCCGAAAGATTGCGGAGAAACATCTTCTAGC 1529  
DB 1206 TTGGCTCTTCGAACACTTTCTTCATCCATACCCCTAAGGATTCGGACAAGCATGCTAGC 1265  
QY 1530 TATACGAAGTGGCTTGACAGAGTTCAGGTATCAAACTGTTTATAAATGCGCGGTTAG 1589  
DB 1266 TAAGCAACAGGACTCACTGTCAGCGAGTGTACATGCGAGGATGTTATTAACCGAGAGTTCG 1325

QY 1590 GCTATGGAAGCCGATGATAGAGAGATGTATGCGGAAATGAACAAGAGAGCTCAATAA 1649  
DB 1326 GTTATGGAACCAATGTTGGAGGAGTGTATCGAGGAAATGAAGGAGCGGCAAGAA 1385  
QY 1650 CA 1651  
DB 1386 CA 1387  
RESULT 12  
US-09-938-842A-1337  
; Sequence 1337, Application US/09938842A  
; Publication No. US20040009476A9  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: SCRIPT1300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 1337  
; LENGTH: 2043  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-1337

Query Match 5.6%; Score 123.6; DB 11; Length 2043;  
Best Local Similarity 69.4%; Pred. No. 6.3e-21;  
Matches 168; Conservative 0; Mismatches 74; Indels 0; Gaps 0;  
QY 1410 CCATCAGATTGGAGACCTCAACGAGGTTTGCTGAGAAATCTGTTTCGGTTCTACGGAA 1469  
DB 1146 CAATAATGCTTGGAGACCTCAACGAGGTTTGCTGAGAAATCTGTTTCGGTTCTACGGAA 1205  
QY 1470 TTGGATGTTCCAAAATCTCTTACCCCTTACCCGAAAGATTGCGGAGAAACATCTTCTAGC 1529  
DB 1206 TTGGCTCTTCGAACACTTTCTTCATCCATACCCCTAAGGATTCGGACAAGCATGCTAGC 1265  
QY 1530 TATACGAAGTGGCTTGACAGAGTTCAGGTATCAAACTGTTTATAAATGCGCGGTTAG 1589  
DB 1266 TAAGCAACAGGACTCACTGTCAGCGAGTGTACATGCGAGGATGTTATTAACCGAGAGTTCG 1325  
QY 1590 GCTATGGAAGCCGATGATAGAGAGATGTATGCGGAAATGAACAAGAGAGCTCAATAA 1649  
DB 1326 GTTATGGAACCAATGTTGGAGGAGATGTATCGAGGAAATGAAGGAGCGGCAAGAA 1385  
QY 1650 CA 1651  
DB 1386 CA 1387

RESULT 13  
US-10-225-066A-335  
; Sequence 335, Application US/10225066A  
; Publication No. US20030226173A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendel Biotechnology, Inc.  
; APPLICANT: Ratcliffe, Oliver  
; APPLICANT: RIECHMANN, Jose Luis  
; APPLICANT: ADAM, Luc J  
; APPLICANT: DUBELL, Arnold T  
; APPLICANT: HEARD, Jacqueline E  
; APPLICANT: PILGRIM, Marsha L

APPLICANT: JIANG, Cai-Zhong  
APPLICANT: REUBER, T. Lynne  
APPLICANT: CREELMAN, Robert A  
APPLICANT: PINEDA, Omaira  
APPLICANT: YU, Guo-Liang  
APPLICANT: BROWN, Pierre E  
TITLE OF INVENTION: Yield-Related Polynucleotides and Polypeptides in Plants  
FILE REFERENCE: MBI0036-2 US/10/225,066A  
CURRENT APPLICATION NUMBER: US/10/225,066A  
CURRENT FILING DATE: 2002-08-09  
PRIOR APPLICATION NUMBER: 09/837,444  
PRIOR FILING DATE: 2001-04-18  
PRIOR APPLICATION NUMBER: 60/310,847  
PRIOR FILING DATE: 2001-08-09  
PRIOR APPLICATION NUMBER: 60/336,049  
PRIOR FILING DATE: 2001-12-05  
PRIOR APPLICATION NUMBER: 60/338,692  
PRIOR FILING DATE: 2001-12-11  
PRIOR APPLICATION NUMBER: 10/171,468  
PRIOR FILING DATE: 2002-06-14  
NUMBER OF SEQ ID NOS: 1122  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 335  
LENGTH: 2385  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-10-225-066A-335

Query Match 6.6%; Score 123.6; DB 13; Length 2385;  
Best Local Similarity 69.4%; Pred. No. 7e-21;  
Matches 168; Conservative 0; Mismatches 74; Indels 0; Gaps 0;  
QY 1410 CCATCAGATTGGAGACCTCAACGAGGTTTGCTCGAGAAATCTGTTTCGGTTCTACGGAA 1469  
Db 1324 CAATATGCTTGGAGACCTCAACGAGGTTTGCTCGAGAAATCTGTTTCGGTTCTACGGAA 1383  
QY 1470 TTGGATGTTCCAAAACCTTCCTTCAACCTTACCCGAAAGATTCGGAGAAACATCTTCTAGC 1529  
Db 1384 TTGGCTCTTCGAACACTTTCTTTCATCCATACCTTAAGGATTCGGACAGCATGCTAGC 1443  
QY 1530 TATACGAGTGGCTTGACAGAGTCAAGTATCAAACTGGTTTATAAATGCGGGTTAG 1589  
Db 1444 TAAGCAACAGAACTCACTCTGAGCGAGTGTGAGCTGGTTTATAAAGCGAGAGTTTCG 1503  
QY 1590 GCTATGGAAGCCGATGATAGAGAGATGATCGGAAATGAACAAGAGAAAGCTCAATAA 1649  
Db 1504 GTTATGGAACCAATGGTGGAGAGATGTACATGAGGAATGAAGGAGCAGGCAAGAA 1563  
QY 1650 CA 1651  
Db 1564 CA 1565

RESULT 14  
US-10-374-780A-2671  
Sequence 2671, Application US/10374780A  
Publication No. US20040019927A1  
GENERAL INFORMATION:  
APPLICANT: Sherman, Bradley K  
APPLICANT: Riechmann, Jose Luis  
APPLICANT: Jiang, Cai-Zhong  
APPLICANT: Heard, Jacqueline E  
APPLICANT: Haake, Volker  
APPLICANT: Creelman, Robert A  
APPLICANT: Ratcliffe, Oliver  
APPLICANT: Adam, Luc J  
APPLICANT: Reuber, T. Lynne  
APPLICANT: Keddie, James  
APPLICANT: Brown, Pierre E  
APPLICANT: Pilgrim, Marsha L  
APPLICANT: Dubell III, Arnold T  
APPLICANT: Pineda, Omaira

TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS  
FILE REFERENCE: MBI-0047 CIP US/10/374,780A  
CURRENT APPLICATION NUMBER: US/10/374,780A  
CURRENT FILING DATE: 2003-02-25  
PRIOR APPLICATION NUMBER: 09/837,944  
PRIOR FILING DATE: 2001-04-18  
PRIOR APPLICATION NUMBER: 60/310,847  
PRIOR FILING DATE: 2001-08-09  
PRIOR APPLICATION NUMBER: 09/934,455  
PRIOR FILING DATE: 2001-08-22  
PRIOR APPLICATION NUMBER: 60/336,049  
PRIOR FILING DATE: 2001-11-19  
PRIOR APPLICATION NUMBER: 60/338,692  
PRIOR FILING DATE: 2001-12-11  
PRIOR APPLICATION NUMBER: 10/171,468  
PRIOR FILING DATE: 2002-06-14  
PRIOR APPLICATION NUMBER: 10/225,066  
PRIOR FILING DATE: 2002-08-09  
PRIOR APPLICATION NUMBER: 10/225,067  
PRIOR FILING DATE: 2002-08-09  
PRIOR APPLICATION NUMBER: 10/225,068  
PRIOR FILING DATE: 2002-08-09  
NUMBER OF SEQ ID NOS: 2906  
SOFTWARE: Patentin version 3.2  
SEQ ID NO 2671  
LENGTH: 2385  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
OTHER INFORMATION: G1589  
US-10-374-780A-2671

Query Match 6.6%; Score 123.6; DB 16; Length 2385;  
Best Local Similarity 69.4%; Pred. No. 7e-21;  
Matches 168; Conservative 0; Mismatches 74; Indels 0; Gaps 0;  
QY 1410 CCATCAGATTGGAGACCTCAACGAGGTTTGCTCGAGAAATCTGTTTCGGTTCTACGGAA 1469  
Db 1324 CAATATGCTTGGAGACCTCAACGAGGTTTGCTCGAGAAATCTGTTTCGGTTCTACGGTT 1383  
QY 1470 TTGGATGTTCCAAAACCTTCCTTCAACCTTACCCGAAAGATTCGGAGAAACATCTTCTAGC 1529  
Db 1384 TTGGCTCTTCGAACACTTTCTTTCATCCATACCTTAAGGATTCGGACAGCATGCTAGC 1443  
QY 1530 TATACGAGTGGCTTGACAGAGTCAAGTATCAAACTGGTTTATAAATGCGGGTTAG 1589  
Db 1444 TAAGCAACAGAACTCACTCTGAGCGAGTGTGAGCTGGTTTATAAAGCGAGAGTTTCG 1503  
QY 1590 GCTATGGAAGCCGATGATAGAGAGATGATCGGAAATGAACAAGAGAAAGCTCAATAA 1649  
Db 1504 GTTATGGAACCAATGGTGGAGAGATGTACATGAGGAATGAAGGAGCAGGCAAGAA 1563  
QY 1650 CA 1651  
Db 1564 CA 1565

RESULT 15  
US-09-922-293-2000  
Sequence 2000, Application US/09922293  
Publication No. US2004012339A1  
GENERAL INFORMATION:  
APPLICANT: Conner, Timothy W.  
APPLICANT: Heck, Gregory R.  
APPLICANT: Liu, Jingdong  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
Transcription in Plants  
FILE REFERENCE: 16517.254  
CURRENT APPLICATION NUMBER: US/09/922,293  
CURRENT FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: US 60/067,000  
PRIOR FILING DATE: 1997-11-24



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----





Db 21 AsnProLeuLeuMetAspPheIleProSerArgGluAspSerThrSerPheSerThrMet 40  
 QY 424 CTTCCATGAATACCATCAGATCCTCTACAAATGGGTGGCTTTGATATTTCAAT 483  
 Db 41 LeuProTrpAsnThrIleArgSerAspProLeuGlnMetGlyGlyPheAspIlePheAsn 60  
 QY 484 TCTATGCTGACTAAACAATACCTATCATCTCTCCACCGTCTATCGATGTTCAAGATAAC 543  
 Db 61 SerMetLeuThrAsnLysThrLeuSerSerProArgSerIleAspValGlnAspAsn 80  
 QY 544 CGCAATGTTGAGTTCATGGCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 603  
 Db 81 ArgAsnValGluPheMetAlaProProHisProProLeuHisProLeuAspHis 100  
 QY 604 TTAAGACACTATGATGATCTCTCAACACATGTGGGTGTTTGAAGCAATAGTGGTTT 663  
 Db 101 LeuArgHisThrAspAspSerSerAsnAsnMetIlePheGluAlaAsnSerGluPhe 120  
 QY 664 CAGGCATTTTCAGGTGTAGTTCGTCCAAAGTGAACCAATGATGCTCTACATTCGGTGAAGAA 723  
 Db 121 GlnAlaPheSerGlyValValGlyProSerGluProMetSerThrPheGlyGluGlu 140  
 QY 724 GATTTCCCGTTCCTAATTCGATTAAGAACAATGAGCTTTCATTCAGTCTTCATCA 783  
 Db 141 AspPheProPheLeuIleSerAsnLysArgAsnAsnGluLeuSerLeuSerLeuAlaSer 160  
 QY 784 GATGTTTCTGATCAATGCTCGAGATAAGTCTTCTGCGAGCTTACAAAGATTAGCCTCAGAG 843  
 Db 161 AspValSerAspGluCysSerGluIleSerLeuCysAlaAlaThrArgLeuAlaSerGlu 180  
 QY 844 CAAGTCTTTCGAGCAGCAAGACATTTCTAATAACGTTGTACTCAAGGTTTCTCTCAA 903  
 Db 181 GlnAlaSerCysSerSerLysAspIleSerAsnAsnValThrGlnGlyPheSerGln 200  
 QY 904 CTTATATTGGCTCAAAATACCTCTCAAGTCTCTCAAGAAATACTATCTCATTTCCGCCCA 963  
 Db 201 LeuIlePheGlySerLysThrLeuHisSerValGlnGluIleLeuSerHisPheAla 220  
 QY 964 TACTCGCTCGATTATTCATCTCAGAGAACCGAGTCCAGGAGTGTAGTTCAGCCTTACT 1023  
 Db 221 TyrSerLeuAspTyrSerSerArgGlyThrGluSerGlyAlaAlaSerSerAlaPheThr 240  
 QY 1024 TCACGTTTTCAGAAATATACTGATGTTCTGATGTTGATTTCTAATACTCCGAGCGGTT 1083  
 Db 241 SerArgPheGluAsnIleThrGluPheLeuAspGlyAspSerAsnAsnSerGluAlaGly 260  
 QY 1084 TTCGGATCTTACATTTCAAAGGAGAGCATTAGAAGCAAAAGAAACCCATCTCTCTGGATCTT 1143  
 Db 261 PheGlySerThrPheGlnArgAlaLeuGluAlaLysThrHisLeuLeuAspLeu 280  
 QY 1144 CTTCAAAATGGTGGATGATGATAGTATTCGCTAGATGATGATCATACGGTTATATCA 1203  
 Db 281 LeuGlnMetValAspAspArgTyrSerHisCysValAspGluIleHisThrValIleSer 300  
 QY 1204 GGTTCATCTGTCACACCGAGTTAGATCCACAGTTTACACACCGGTTTCCCTCCAAACC 1263  
 Db 301 AlaPheHisAlaAlaThrGluLeuAspProGlnLeuHisThrArgPheAlaLeuGlnThr 320  
 QY 1264 GTTTCCTCTTTATACAAGAACCTGAGAGAGAGATCTCTCAAGAGATTAATCTCTATGGGA 1323  
 Db 321 ValSerPheLeuTyrLysAsnLeuArgGluArgIleCysLysLysIleSerMetGly 340  
 QY 1324 TCTGTATTTCGAGAGGCAACACAGACTCAAGAAACCTCTATGTTCCACAGCATTCG 1383  
 Db 341 SerValLeuGluArgGlyLysAspLysThrGlnGluThrSerMetPheHisGlnHisCys 360  
 QY 1384 CTTCTTCAGCAGCTGAACGAAAGAACCATCATGATTTTCGAGACCTCAACGAGGTTGCTT 1443  
 Db 361 LeuLeuGlnGlnLeuLysArgLysAsnHisGlnIleTrpArgProGlnArgGlyLeuPro 380  
 QY 1444 GAGAAATCTGTTTCGGTCTACGGATTGGATGTTCCAAACTCTCTCACCCCTTACCCG 1503  
 Db 381 GluLysSerValSerValLeuArgAsnTrpMetPheGlnAsnPheLeuHisProTyrPro 400

QY 1504 AAGATTCCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAGAAGTCAAGTATCA 1563  
 Db 401 LysAspSerGluLysHisLeuLeuAlaIleArgSerGlyLeuThrArgSerGlnValSer 420  
 QY 1564 AACTGCTTTTATAAATCGCGGGTGTAGCTATGAGACCCGATGATGAAGAGATGTATGCG 1623  
 Db 421 AsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetIleGluGluMetTyrAla 440  
 QY 1624 GAAATGAACAAGAGGAGCTCAATAACAGTCACATTCACATTCACCCCAACGACCACTCTTCA 1683  
 Db 441 GluMetAsnLysArgLysLeuAsnAsnSerHisIleGlnProAsnGlyProThrLeuArg 460  
 QY 1684 ATGCCAAATCTCTTATGATGAGCCCAAGCAATGCATAAA 1722  
 Db 461 MetProLysSerValMetMetSerGlnAlaMetHisLys 473  
 RESULT 2  
 H84774  
 probable homeodomain transcription factor [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: H84774  
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; A.  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.B.; Umayam, L.; Tallon, L.  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 A:Accession: H84774  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-680 <STO>  
 A:Cross-references: GB:AE002093; NID:g4510375; PIDN:AAD21463.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2G35940  
 A:Map position: 2  
 Alignment Scores:  
 Pred. No.: 2,65e-29 Length: 680  
 Score: 472.00 Matches: 152  
 Percent Similarity: 42.01% Conservative: 74  
 Best Local Similarity: 28.25% Mismatches: 193  
 Query Match: 14.37% Indels: 119  
 DB: 13 Gaps: 13  
 US-09-423-575-1 (1-1886) x H84774 (1-680)  
 QY 283 ATAAAGTTTCATAGAAACCAATGGACAAACAACAACAACACTTTTAGTTCTCTG 342  
 Db 32 ValGlnTyrThrGlnGlnAspAsnAspSerAsnAsnAsnAsnAsnSerAsnAsnSerAsn 51  
 QY 343 GAT-----AATGTCATGACTAACCAAAATCCCTCTCTCATGGATTTTACCTTCA 393  
 Db 52 AsnAsnAsnThrAsnThrAsnThrAsnAsnAsnAsnSerSerPheValPheLeuAspSer 71  
 QY 394 AGAGAAGATTCACATTCATTCACCAATGCTTCCATGGAATACCATCAGATCAGATCCT 453  
 Db 72 HisAlaPro-GlnProAsnAlaSerGlnGlnPheValGlyIleProLeu----- 87  
 QY 454 CTACAAATGGGTGGCTTTCATATTTCAATTCATGCTGACTAACAAATACCTATCATCT 513  
 Db 88 -----SerGlyHis-- 90  
 QY 514 TCTCCAGGCTCTATCGATGTTCAAGATAACCGCAATGTTGAGTTTCATGCTCTCTCTCT 573  
 Db 91 GluAlaAlaSerIleThrAlaAlaAspAsnIleSerValLeuHisGlyTyrProProArg 110  
 QY 574 CATCCCTCCCTCCATCTCTTGGATCATTTTAAGACACTATGATGATTCCTCAACAAAC 633  
 Db 111 ValGlnTyrSerLeuTyrGly-----SerHisGlnValAspProThrHisGln 126  
 QY 634 ATGTGGGTTTTCAGCAACAAATAGTAGTTTCAGGCATTTTCAGGTGTAGTTGGTCCAAGT 693



```

QY 687 TCCAGTCAACCAATGATCTCTACATTCGGTGAAGAAGATTTCCTGTTTCTTAATTCGAA 746
Db 117 -----SerThrMetSerAspGlu-----
QY 747 TAAAGAAACATGAGCTTTTCAATGAGCTTCGATCAGATGTTTCTGATGATGCTCGGA 806
Db 123 -----AsnGlySerLeuSerValHisGlnHisSerAspGlnLeuPr 139
QY 807 GATAGTCTTTGTGAGCTACAAGATTAGCTCAGAGCAAGCTTCTTGACAGCAAGA 866
Db 139 oSerSerVal-----TyrAsnAs 145
QY 867 CATTTCTAATAACGTTGTT-----ACTCAAGTTTCTC 899
Db 145 nAsnGlyAsnAsnGlyValGlyPheTyrAsnAsnTyrArgTyrGluThrSerGlyPheVa 165
QY 900 TCACTTATATTGGCTCAAAATACCTTCACCTGTTCACAAATACATCTCATCTTCGC 959
Db 165 lSerSerValLeuArgSerA-glyLeuLysProThrGlnGlnLeuLeuAspGluValVa 185
QY 960 CGCATCTCGCTCGATTAT-----TCATCTCGAGGAACCGA 995
Db 185 lSerValArgLysAspLeuLysLeuGlyAsnLysLysMetLysAsnAspLysGlyGlnAs 205
QY 996 GTCAGAGCTGCTAGTTTCAGCTTACTTCACGTTTTCAGATATATACTGATTTCTTGA 1055
Db 205 pPheHisAsnGlySerSer-----AspAsnIleThrGlu-----As 217
QY 1056 TGGTCAATCTAATAACTCGGAGCGGGTTTCGGATCTACATTTCAAAGGAGGACATTAGA 1115
Db 217 pAspLysSerGlnSerGlnGluLeu-----SerProSerGluArgGlnGluLeuGl 234
QY 1116 AGCAAGAAACCCATCTCTGGATCTTCTCAATGGTGTGATGATGATATAGTATG 1175
Db 234 nSerLysLysSerLysLeuLeuThrMetValAspGluValAspLysArgTyrAsnGlnTy 254
QY 1176 COTAGATGAGATTATACCGTTATATACAGCTTCATCTGCTCAACCGAGTTAGAT---CC 1232
Db 254 rHisGlnMetGluAlaLeuAlaSerPheGluMetValThrGlyLeuGlyAlaAl 274
QY 1233 ACAGTTACACACCCGGTTGGCTTCCCAACCGCTTTCCTTATACAAAGACCTTGAGAG 1292
Db 274 aLysProTyrThrSerValAlaLeuAsnArgIleSerArgHisPheArgCysLeuArgAs 294
QY 1293 GAGATCTGCAAGAGATA---ATCTCTATGGATCTGTATTGAGAGAGCGCAAGACAA 1349
Db 294 pAlaIleLysGluGlnIleGlnValIleArgGlyLysLeuGlyGluArgGluThrSerAs 314
QY 1350 GACTCAAGAAACCTCTATGTTCCACCAAGCAATGCTTCTTCAGAGCTGAAACGAAAGAA 1409
Db 314 pGluGlnGlyLysArgIleProArgLeuArgTyrLeuAspGlnArgLeuArgGlnGlnAr 334
QY 1410 C-----CATCAG-----ATTGGAGACTCAACGAGTTTGC 1442
Db 334 gAlaLeuHisGlnGlnLeuGlyMetValArgProAlaTrpArgProGlnArgGlyLeuPr 354
QY 1443 TGAGAAATCTGTTTCGGTTCTACGGAATTTGGATGTTCCAAATCTCTTCCACCTTACCC 1502
Db 354 oGluAsnSerValSerIleLeuArgAlaTrpLeuPheGluHisPheLeuHisProTyrPr 374
QY 1503 GAAAGATTCGAGAAACATCTCTAGCTATACGAAGTGGCTTGACAGAGCTAGGTATC 1562
Db 374 oLysGluSerGluLysIleMetLeuSerLysGlnThrGlyLeuSerLysAsnGlnValAl 394
QY 1563 AAACCTGGTTTAAATCGCGGTTAGCTATGAAGCCGATGATAGAAGAGATGTATGC 1622
Db 394 aAsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetIleGluGluMetTyrLy 414
QY 1623 GGAA-----ATGACAAGAGGAAGCTCAATAACAGTCAATTCACCCACCGAGCC 1673
Db 414 sGluGluPheGlyGluSerAlaGluLeuLeuSerAsnSer-----AsnGlnAs 430
QY 1674 AACTCTTCGAATGCCAAATCTGTTATGATGAGCCAGCAATGCATATAA----- 1722
Db 430 pThrLysLysMetGlnGluThr-----SerGlnLeuLysHisGluAspSerSerSe 447
QY 1723 -----TAAGACAACAATTTGTTTACCAACTTTGTGATAATTA 1760
Db 447 rSerGlnGlnGlnAsnGlnGlyAsnAsnAsnAsnIleProTyrThrSerAsp-AlaG 467
QY 1761 GGCAATTTGCTACTCTATGATTCGCCAAACCTTAAACCATGTACGACT 1807
Db 467 luGlnAsnLeuValPheAlaAspProLysProAspArgAlaThrThr 482
RESULT 4
H86329
F699.25 protein - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
C/Accession: H86329
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A/Reference number: A86141; MUID:21016719; PMID:11130712
A/Accession: H86329
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1538 <STO>
A/Cross-references: GB:AE005172; NID:g10086497; PIDN:AAG12557.1; GSPDB:GN00141
C/Genetics:
A/Map position: 1
Alignment Scores:
Pred. No.: 1,22e-25 Length: 538
Score: 425.50 Matches: 119
Percent Similarity: 43.04% Conservative: 59
Best Local Similarity: 32.78% Mismatches: 154
Query Match: 12.96% Indels: 31
DB: 2 Gaps: 9
US-09-423-575-1 (1-1886) x H86329 (1-538)
QY 746 ATAAAGAAACATGAGCTTTTCAATGAGCTTCGATCAGATGTTTCTGATGATGCTCGG 805
Db 111 ValLysGluThrSerProPheHis-----ValAspGluMet--SerValLysSerLysG 128
QY 806 AGATAAGTCTTTGTGCGAGCTACAGATTAGCTTCAGAGCAAGCTTCTTCGACAGCAAG 865
Db 128 luMetIleLeuLeuGlyGlnSerAspProSerSerGlyTyrAlaGlyAsnGlyLysG 148
QY 866 ACATTTCTTAATAAGCTT-----GTTACTCAGGTTTCTCTCACTTATATTG 913
Db 148 lYPheTyrAsnAsnTyrArgTyrAsnGluThrSerGlyLysPheMetSerSerValLeuA 168
QY 914 GCTCAAAATACCTTCACCTCTGTTCAAGAAATACATCTCATTTCCGCCGACTACTCGCTCG 973
Db 168 rSerArgTyrLeuLysProAlaGlnAsnLeuLeuAspGluValSerValLysLysG 188
QY 974 ATTATTCTCTCGAGAACCCGAGTCAGGAGCTGCTAGT-----TCAGCTTTTACTTCA 1027
Db 188 luLeuAsnGlnMetGlyLysLysMetLysValAsnAspPheAsnSerGlySerLysG 208
QY 1028 GTTTTCAGATATACTGAGTTTCTTGATGGTATTCTTAATAACTCGGAGCGGGTTTCG 1087
Db 208 luIleGluGlyGlyGlyGluLeuSerSerAspAsnGlyLysSerIleGluLeu- 227
QY 1088 GATCTACATTTCAAAGGAGGACATTAGAGCAAGAAACCATCTCTTGATCTTCTTCTTC 1147
Db 1147 -----

```

[illegible]

Alignment Scores:	
Pred. No.:	2 61e-25
Score:	421.50
Percent Similarity:	41.12%
Length:	160
Matches:	90
Conservative:	90

Best Local Similarity:	26.32%	Mismatches:	168
Query Match:	12.83%	Indels:	190
DB:	2	Gaps:	26

  

US-09-423-575-1 (1-1886) x D85435 (1-638)						
Qy	289	TTTCATAGAACCCCAATGGACAACAC	----	315		
Db			:::			
Db	17	PheHisAsnAsnSerMetSerGlnAspTyrHisHisHisHisHisAsnGlnHisGln	36			
Qy	316	-----AACAAACAACACTTTTAGTTCTCTGGAT	-----AATGTCAATGACT	357		
Db		::    :				
Db	37	GlyGlyLePheAsnPheSerAsnGlyPheAspArgSerAspSerProAsnLeuThr	56			
Qy	358	AACCAAATCTCT	-----CTTCTCATGGAT	381		
Db		::				
Db	57	GlnGlnGlnGlnGluHisGlnArgValGluMetAspGluGluSerSerValAlaGlyGly	76			
Qy	382	TTTATACCTTCAAGAGAAGATCAACTTCATTCTCAACAATGCTTCATCGAATACCATC	441			
Db		::				
Db	77	ArgIleProValTyGluSerAlaGlyMetLeuSerGluMetPheAsnPhe	93			
Qy	442	AGATCAGATCCTCTCAAAATGGTGGC	-----TTTGATATTTTTCAATTTCTATGCTG	492		
Db						
Db	94	-----ProGlySerSerGlyGlyGlyArgAspLeuAspLeuGlyGlnSerPheArg	110			
Qy	493	ACTACAAATACTATTATCTCTCTCCACGGTCTATCGATGTTCAAGATAACCGCATGTT	552			
Db		::				
Db	111	SerAsnArgGlnLeuLeuGlu	-----GluGlnHisGlnAsnIle	123		
Qy	553	GAGTTTCATG	-----	561		
Db						
Db	124	ProAlaMetAsnAlaThrAspSerAlaThrAlaAlaAlaMetGlnLeuPheLeu	143			
Qy	562	-----GCTCTCTCTCTCATCTCTCTCCA	-----	585		
Db						
Db	144	MetAsnProProProGlnProProSerProSerSerThrThrSerProAsgser	163			
Qy	586	-----CTTCAT	-----CCTTTGGATCATTTTAAGACAC	612		
Db						
Db	164	HisHisAsnSerSerThrLeuHisMetLeuLeuProSerProSerThrAsnThrThrHis	183			
Qy	613	TATGATGATCTCTCAACAACATG	-----	636		
Db		::: ::    ::	-----			
Db	184	HisGlnAsnTyThrAsnHisMetSerMethisGlnLeuProHisGlnHisGlnGln	203			
Qy	637	-----TGGGGTTTGAACAAAATGACGTTCAGGCATTTTCAGGTGTAGTTGGT	687			
Db		::				
Db	204	IleSerThrTrpGlnSerSerProAspHisHisHisHisAsn	219			
Qy	688	CCAAGTGAACCAATGATCTCATCTCGGTCAAGACATTCCTCGTTTCTTAATTTCCGAT	747			
Db		::				
Db	220	---SerGlnThrGluIleGlyThrValHisValGluAsn	234			
Qy	748	AAAAAGAACATGACTTTTCATTGAGCTTCGCATCATGATT	792			
Db		::				
Db	235	HisGlyGlyGlnGlyLeuSerLeuSerLeuSerSerLeuGluAlaAlaLysAla	254			
Qy	793	GATGAATGTCGGACATAAGCTTTGTGACAGTACACAGATTACCTCA	840			
Db		::				
Db	255	GluGluTyArgAsnIleTyTyGlyAlaAsnSerSerAsnAlaSerProHisHisGln	274			
Qy	841	-----GACCAAGCTTCTTGACAGCAAGACATTTCTTAATACGTTGTACT	888			
Db		::				
Db	275	TyrAsnGlnPheLysThrLeuLeuAlaAsnSerSerGlnHisHisGlnValLeuAsn	294			
Qy	889	CAAGGTTCTCT	-----CAACTATATTT	912		
Db						
Db	295	GlnPheArgSerSerProAlaAlaSerSerSerMetAlaAlaValAsnIleLeuArg	314			
Qy	913	GGCTCAAAATACCTTCACTCTGTTTCAAGAAATACTATCTCATTTGCCCGCATCTGCCTC	972			
Db		::				
Db	315	AsnSerArgTyThrThrAlaAlaGlnGluLeuGluGluPheCysser--ValGlyA	334			

A;Cross-references: GB:AE002093; NID:gy454389; PIDN:AAD22299.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: Ac2g16400  
A;Map position: 2

Alignment Scores:

Pred.	No.:		Length:	482
Argument Scores:				
Score:		3.92e-25	Matches:	118
Percent Similarity:		47.0%	Conservative:	53
Best Local Similarity:		32.5%	Mismatches:	120
Query Match:		12.76%	Indels:	72
Gaps:		2	Gaps:	11

US-09-423-575-1 (1-1886) x H84539 (1-482)

619	QY	GATTCTCTCAACAAACATGTGGGGTTTTTGAAGCAAAATAGTGCAGTTCAGGCATTTTCAGGT	678
37	Db	AspSerThrAsnAsnVal-----SerProSerLysGluIleGlnValLeuSerSer	53
679	QY	GTAGTTGGTCCCAAGTGAACCAATGATG-----TCTACATTCGGTGAAGAA	723
54	Db	LeuGlyGlyValSerGlnMetValGluIleGlnAspSerClySerIrpArgAspGlnGlu	73
724	QY	GATTTCCCGTTTCTAATTCGGAATAAAGAAACAAT-----GAGCTTTCATTTCAGT	774
74	Db	Asp-----AsnAspArgAsnArgPheProValMetArgArgLeuGly	87
775	QY	CTTGTCATCAGATGTTTCTGATGAATGCTCGGAGATAAGTCTTTGTGCGACTACAAGATTA	834
88	Db	LeuSerSerGlnIleGlu-----	93
835	QY	GCCTCAGACGAAGCTTCTTCGACGACGAAGAAGACATTTCTTAATTAAC-----GTT	882
94	Db	-----ThrSerArgGlyAsnAsnAsnGluTyrAlaThrGln	106
883	QY	GTTTACTCAAGTTTCTCTCAACTATATTGGCTCAAAATACCTTCACCTCTGTTCAGAA	942
107	Db	ValValSerGlyPheThrArgThrIleHisAsnSerLysTyrlLeuLysAlaGlnGlu	126
943	QY	ATACTATCTCATTTCCGCGGATACTCGCTCGATATTATCTCTCGAGAAACCGAGTCAGGA	1002
127	Db	LeuLeuAspGluThrValAsnValLys-----	135
1003	QY	GCTGCTAGTTCACGCTTTTACTTCCAGCTTTTGGAAATATAACTCAGTTTCTTGTGCTGTGAT	1062
136	Db	LysAlaLeuLysGlnPheGlnProGluGlyAspLysIleAsnGluValLysGluLysAsn	155
1063	QY	TCTAATAACTCGGAGCGGGTTTCGGATCTACATTTCAAAGAGAGACGATTAGAACGAAG	1122
156	Db	LeuGlnThrAsnThrAlaGluIleProGlnAla--GluArgGlnGluLeuGlnSerLys	174
1123	QY	AAAACCATCTCTGGATCTTCTTCAAATGGTGGATGATCGATATAGTCATTCCGTAGAT	1182
175	Db	LeuSerLysLeuLeuSerIleLeuAspGluValAspArgAsnTyrlLysGlnTyrlHis	194
1183	QY	GAGATTCATACGGTTATTATCAGCGTTCCATGCTGCAACCGAGTTAGAT---CCACAGTTA	1239
195	Db	GlnMetGlnIleValValSerSerPheAspValIleAlaGlyCysGlyAlaAlaLysPro	214
1240	QY	CACACCGGTTTCCCTCCAAACCGTTTCTTCTTATACAAAGAACCTTGAGAGAGAATC	1299
215	Db	TyrThrAlaLeuAlaLeuGlnThrIleSerArgHisPheArgCysLeuArgAspAlaIle	234
1300	QY	TGCAAGAAGATAATC-----TCT	1317
235	Db	SerGlyGlnIleLeuValIleArgLysSerLeuGlyGluGlnAspGlySerAspGly	254
1318	QY	ATGGGATCTGTATTGGAGAGAGCAAGAACACACACTCAAGAAACCTCTATGTTCCACCAG	1377
255	Db	ArgGlyValGlyIleSerArgLeuArgAsnValAspGlnIleVal-----ArgGln	271
1378	QY	CATTGGCTTCTTCAGCAGCTGTAACGAAAGAACCATCAGATTTGGAGACCTCAACGAGGT	1437

973 GATTATTCACTCGAGAACCGAGTCCGAGAGCTGTAGTTCACGCTTTTACTTTCACGTTTT 1032

334 rgglyPheLeuLysLysAsnLysLeu-GlyAsnSerAsnProAsnThrCysGlyGly 353

1033 GAGAATATAACTGAGTTTCTTGATGTGTATCTTAATAACTCGGAGCGGTTCGCA--- 1089

354 AspGly-----GlyGlySerSerProSerSerAlaGlyAlaAsnLys 367

1090 -----TCTCATTTCAAAGAGAGAGCATTAAGCAAGAAGAAACCCATCTC 1134

368 GluHisProProLeuSerAlaSerAspArgGluHisGlnArgArgLysValLysLeu 387

1135 TTGGATCTTCTTCAAATGGGATGATCGATATAGTCAATTCGCTAGATGAGATTTCATACG 1194

388 LeuThrMetLeuGluGluValAspArgArgTyrAsnHisTyrCysGluGlnMetGlnMet 407

1195 GTTATATCAACGCTC-----CATGCTCCAAACCGAGTTAGATCCACAGTTA 1239

408 ValValAsnSerPheAspIleValMetGlyHisGlyAlaAla-----LeuPro 423

1240 CACACCCGGTTTCCCTCCAAACCGTTTCTTCTTATACAAGAACCTCGAGAGAGAACTC 1299

424 TyrThrAlaLeuAlaGlnLysAlaMetSerArgHisPheArgCysLeuLysAspAlaVal 443

1300 TGAAGAAGAGATA-----ATCTCT 1317

444 AlaAlaGlnLeuLysGlnSerCysGluLeuLeuGlyAspLysAspAlaAlaGlyIleSer 463

1318 ATGGGATCTCTATTGGAGAGAGCAAGACAACTCAAGAAACCTCTATGTTCCACCAG 1377

464 ---SerSerGlyLeuThrLysGly-----GluThrProArgLeuArg--- 476

1378 CATTCGCTTCTTCAGCAGCTGAAACCAAGAAAC-----CATCAGATT----- 1419

477 -----LeuLeuGluGlnSerLeuArgGluAsnArgAlaPheHisGlnMetGlyMetGlu 494

1420 -----TGGAGACCTCAACAGGTTTCCTTGAGAAATCTGTTTCGGTTTACCGAAT 1470

495 GlnGluAlaTrpArgProGlnArgGlyLeuProGluArgSerValAsnIleLeuArgAla 514

1471 TGGATGTTCCAAACTTCCTTCCACCTTACCCCAAGATTTCGGAGAAACATCTTTCAGCT 1530

515 TrpLeuPheGluHisPheLeuHisProTyrProSerAspAlaAspLysHisLeuLeuAla 534

1531 ATACGAAGTCGGTTGCACAGAAGTCAGGTATCAAACTGGTTTATAATCGCGGGTTAGG 1590

535 ArgGlnThrGlyLeuSerArgAsnGlnValSerAsnTrpPheIleAsnAlaArgValArg 554

1591 CTATGGAAGCCGATGATAGAGAAGATGTATGCGGAATGAACAGAGGAAGCTCAATAAC 1650

555 LeuTrpLysProMetValGluGluMetTyrGlnGlnLysSerLysGluArgGluArgGlu 574

1651 AGTCACATTCACCCCAAC 1668

575 GluGluLeuGluGluAsn 580

RESULT 6

H84539

Probable homeodomain transcription factor [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C/Accession: H84539

R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999

A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A/Reference number: A84420; MUID:20083487; PMID:10617197

A/Accession: H84539

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-482 <STO>



Db 272 GlnArgAlaLeuGlnArgLeuGlyValMetGlnProHisThrTrpArgProGlnArgGly 291

Qy 1438 TTGGCTCGAGAAATCTGTTTCGGTTCTACGGAATTTGGATGTTCCAAAACCTTCCTTCCACCT 1497

Db 292 LeuProAspSerValLeuValLeuAlaTrpLeuPheGluHisPheLeuHisPro 311

Qy 1498 TACCCGAAAGNTTCGGAGAAACATCTCTTACGTATACGAAGTCGCTTGACAGAAATCAG 1557

Db 312 TyrProLysAspSerAspLysIleMetLeuAlaArgGlnThrGlyLeuSerArgGlyGln 331

Qy 1558 GTATCAAACTGGTTTATAAATGCGCGGTTAGGCTATGGAAGCGCATGTAGAGAAGAGATG 1617

Db 332 ValSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetValGluGluMet 351

Qy 1618 TATCGGAA 1626

Db 352 TyrLysGlu 354

RESULT 7

C84670

C:Probable homeodomain transcription factor [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: C84670

M.; Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:2C083487; PMID:10617197

A:Accession: C84670

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-431 <SPT>

A:Cross-references: GB:AE002093; NID:g6598540; PIDN:AAF18625.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g27220

A:Map position: 2

<b>Alignment Scores:</b>					
Pred. No.:	5,04e-25	Length:	431		
Score:	417.50	Matches:	102		
Percent Similarity:	50.83%	Conservative:	51		
Best Local Similarity:	33.89%	Mismatches:	91		
Query Match:	12.71%	Indels:	57		
DB:	2	Gaps:	9		
 US-09-423-S75-1 (1-1886) x C84670 (1-431)					
QY	841	GAGCAGCTTCTTCGACGACAAGACATTTCTAATAACGTTCCTTA	CTTACTCAAGTTCCTCT	900	
Db	60	GLnGLnCLyAsnSerIleSerThrPheSerASnGLyValPheaArgAlaLeuAla	79		
QY	901	CAACTTATATTGGCTCAAATACTTCATCTGTTCAGAATACTA	-----	948	
Db	80	Profile-----TyrLeuLysAlaAlaGLnGLuLeuLysnGLulleVal	94		
QY	949	-----TCTCATTTCCGCCGANACTCGCTCGATTATTCATCTCGAGAAC	993		
Db	95	AsnValGLyAsnGLyserHisGLyAlaLysGLnGLuArgProvalSerLysGlusThr	114		
QY	994	GAGTCAGAGCTGCTAGTTCCAGCCTTTACTTCACGTTTTGAGAATAAACTCAGITTC	TCT	1053	
Db	115	IleTyrgly-----ValGLuasprile	121		
QY	1054	GATGGTGATCTTAATACTCGAGGGGGTTTCGGATCTACATTTCAAGGAGAGCATTA	1113		
Db	122	AsnGLyGLyTyrLysProGLyValAla-	-----AlaLeu	132	
QY	1114	GAAGCAAAGAAAACCCATCTCTTGGATCTCTTCRAATGGTGCATCATGATATAGTCAT	1173		
Db	133	GLnMetLysLysalylsueirieserMetGLygluMetValGLuGLnArgTyrLysGLn	152		

QY	1174	TGCGTAGATGAGATTTCATACGGTTTATATCATACGGCTTCCATGCTCCAAACCGAGTTA--GAT	1230
Db	153	TyrHisaspGlnMetGlnThrIleSerSerPheGluclnAlaLagIleuGlySer	172
QY	1331	CCACAGTTACACACCCGGTTGGCCTCCAAACCGTTTCCTTCTATACAGAACCTGAGA	1290
Db	173	AlaAsnSerTyrThrHisMetAlaLeuGlnThrIleSerLysGlnPheArgAlaValLys	192
QY	1291	GAGAGAATCTCGAAGAACATAAATCTCTATGGGATCTGTATTTGGAGAGAGCAAGCAAG	1350
Db	193	AspMetIleSerLeuGlnIleLysGlnIleAsnLysLeuLeu-----GlyGlnLysGlu	210
QY	1351	ACTCAAGAAACC-----TCTATGTTCCACGAGATTGCCTTCTTCACGAG	1395
Db	211	PheAspGlnGlnLeuLysLysLeuGlyLysMetAlaHisHis-----	225
QY	1396	CTGAACGAAAGAACCATCAGATTGGAGACCTCAACGAGTTTGCTCGAGAATCTGTT	1455
Db	226	-----SerAsnAlaTPrpGlnArgGlyLeuProGlnLysAlaVal	240
QY	1456	TCGTTTTCACGGAATTGGATGTTTCCAAAACCTTCCTTCAACCCTTACCAGAAATTCGGAG	1515
Db	241	SerValLeuArgSerTrpLeuPheGluHisPheLeuHisProTyrProArgAspLeuAsp	260
QY	1516	AAACATCTTCTAGCTATACCAAGTCGCTTGACAGAAGTCAGGTATCAAACTGGTTTATA	1575
Db	261	LysValMetLeuAlaLysGlnThrGlyLeuThrLysSerGlnValSerAsnTrpPheIle	280
QY	1576	AATCGCGGTTTAGGCTATCGAAGCCGATGATAGAAGATGTATCGGAA-----	1626
Db	281	AsnAlaArgValArgMetTrpLysProLeuValGluGluLeuTyrSerGluMetAsp	300
QY	1627	ATGACACAGAGGAAGCTCAATACAGTCACATCAACCCAAACCGACCAACTCTTCAATG	1686
Db	301	IleGluGluSerArgLysGlySerAspArgTyrSerThrLysGlySerSerLysGln	320
QY	1687	CCA 1689	
Db	321	Pro 321	
RESULT 8			
T05281			
probable homeobox protein T4L20.190 - Arabidopsis thaliana			
C:Species: Arabidopsis thaliana (mouse-ear cross)			
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999			
C:Accession: T05281			
R:Bevan, M.; Terry, N.; Ardiles, W.; Buyschaert, C.; Dasseville, R.; De Clerck, H.W.; Mayer, K.F.X.; Schueller, C.			
submitted to the Protein Sequence Database, September 1998			
A:Reference number: Z15406			
A:Accession: T05281			
A:Molecule type: DNA			
A:Residues: 1-532 <BEV>			
A:Cross-references: EMBL:AL023094			
A:Experimental source: cultivar Columbia; BAC clone T4L20			
C:Genetics:			
A:Map position: 4			
A:Introns: 217/3; 340/2; 360/3			
A:Note: T4L20.190			
Alignment Scores:			
Pred. No.:		5,24e-25	Length: 532
Score:		417.50	Matches: 134
Percent Similarity:		44.52%	Conservative: 61
Best Local Similarity:		30.59%	Mismatches: 159
Query Match:		12.71%	Indels: 85
DB:		2	Gaps: 12
US-09-423-575-1 (1-1886) x T05281 (1-532)			
QY	481	AATCTATGCTGACTAACAAATCTTATCACTTCTCCACGGCTCTATCGATTTCAAGAT	540
Db	13	AspSerMetIleGlnAsnAlaIleValSerTyrSerGluSerAlaGly-----	29



Db 146 GlnGlnGlnGlyTyrSerGlyAsnLysSerThrGlnHisGlnAsnLeuGlnHisThrGln 165  
QY 829 AGATTAGCCTCAGACCAAGCTCTTCAGACAGACAGACATTTCTAATAACGTTGTACT 888  
Db 166 MetMetMetMetMetMetMetMetMetMetMetMetMetMetMetMetMetMetMet 185  
QY 889 CAAGGTTCTCTCAACTTATATTGGC---TCAAAATACCTTCACCTCTGTTCAAGAAATA 945  
Db 186 HisAsnHisHisGlnPheGlnGlySerSerLysThrLeuSerProAlaGlnGluLeu 205  
QY 946 CTATCTCATTTCCGCCACTCGCTCGATTATTCATCT----- 984  
Db 206 LeuSerGluPheCysSerLeuGlyValLysGluSerAspGluValMetMetMetLys 225  
QY 985 ---CGAGGACACCGAGTCAGAGCTGTAGTTACGCTTT---ACTTCAGCTTTTTCAGATT 1038  
Db 226 HisLysLysLysGlnGlyGlnGlnGlnGluTyrAspThrSerHisHisSerAsn 245  
QY 1039 ATAACTGAGTTCTTTCATGGTGTATCTTAATACTCGAGCGCGGTTTCGATCT----- 1092  
Db 246 -----AsnAspGlnHisAspGlnSerAlaThrThrSerSerLysLys 259  
QY 1093 -----ACATTTCAAAGGAGAGCATTAGAACCAAGAAACCCCATCTC 1134  
Db 260 HisValProLeuHisSerLeuGluPheMetGluLeuGlnLysArgLysAlaLysLeu 279  
QY 1135 TTGGATCTCTTCAAATGGTGATGATCGATATAGTCATTGCGTAGATGATTCATAG 1194  
Db 280 LeuSerMetLeuGlnGluLeuLysArgArgTyrGlyHisTyrArgGluGlnMetArgVal 299  
QY 1195 GTTATATACGCTTCCATCGTCAACCGAGTAT---GATCCACAGTTACACACCGGTTT 1251  
Db 300 AlaAlaAlaAlaPheGluAlaValGlyLeuGlyAlaGluLysThrAlaLeu 319  
QY 1252 GCCTCCAAACCGTTTCTCTTATACAAAGACCTCGAGAGAGATCTGCAAGAGATA 1311  
Db 320 AlaserArgAlaMetSerArgHisPheArgCysLeuLysAspGlyLeuValGlyGlnIle 339  
QY 1312 ATCTCTATGGATCTGTATTG---GAGAGAGGCAAGAC----- 1347  
Db 340 GlnAlaThrSerGlnAlaLeuGlyGluArgGluGluAspAsnArgAlaValSerIleAla 359  
QY 1348 -----AAGACTCAAGAAACCTCTATGTTCCACAGCATTGCTTCTTCAGAGCTG 1398  
Db 360 AlaArgGlyGluThrProArgLeuArgLeuLeuAspGlnAlaLeuArgGlnLysSer 379  
QY 1399 AAGCAAGAAACCATCAGATT-----TGAGACCTCAACAGAGTTTGCTGAG 1446  
Db 380 TyrArgGlnMetThrLeuValAlaAspAlaHisProTyrArgProGlnArgGlyLeuProGlu 399  
QY 1447 AAATCTGTTTGGCTTCTACGGAATTCGATGTTCCAAACCTTCTCACCTTACCGGAA 1506  
Db 400 ArgAlaValThrThrLeuArgAlaTyrLeuPheGluHisPheLeuHisProTyrProser 419  
QY 1507 GATTCGGAGAAACATTTTAGCTATACGAAGTGGCTTCACAAGAGTCAGGATCAAAAC 1566  
Db 420 AspValAspLysHisIleLeuAlaArgGlnThrGlyLeuSerArgSerGlnValSerAsn 439  
QY 1567 TGGTTTATAATGCGCGGTTAGGTTATGAGCGCGATGATGAGAGAGATGATCGGAA 1625  
Db 440 TrpPheIleAsnAlaArgValArgLeuTyrPheProMetIleGluMetTyrCysGlu 459  
QY 1627 ATGAACAAGAGGAAGTCAATACAGTCACTTCAACCA----- 1666  
Db 460 Glu-ThrArgSerGluGlnMetGluIleThr---AsnProMetMetIleAspThrLysPr 478  
QY 1667 ----ACGGACCAACTCTTCGATGCCAA 1690  
Db 478 cAspProAspGlnLeuIleArgValGlu 487

RESULT 10  
T02415

probable homeodomain transcription factor [imported] - Arabidopsis thaliana

N:Alternate names: probable BEL1-like homeotic protein  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 16-Feb-2001  
R:Accession: T02415; F84628  
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, submitted to the EMBL Data Library, May 1998  
A:Description: Arabidopsis thaliana chromosome II BAC F27L4 genomic sequence.  
A:Reference number: 214658  
A:Accession: T02415  
A:Status: translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-627 <R0U>  
A:Cross-references: EMBL:AC004482; NID:g3152602; PIDN:AAC17087.1; PID:g3152608  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, Nature 402, 763-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: F84628  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-627 <STO>  
A:Cross-references: GB:AB020093; NID:g3152608; PIDN:AAC17087.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: F27L4.6; At2g23760  
A:Map position: 2  
A:Introns: 319/3; 450/2; 470/3

Alignment Scores:  
Pred. No.: 3,05e-24 Length: 627  
Score: 408.00 Matches: 142  
Percent Similarity: 41.48% Conservative: 82  
Best Local Similarity: 26.30% Mismatches: 141  
Query Match: 12.42% Indels: 176  
Gaps: 2

US-09-423-575-1 (1-1886) x T02415 (1-627)

QY 310 AACAAACAACAACAACACTTTTAGTTCTGTGATAATGTCATGACTAACCAAAATCCT 369  
Db 93 AsnAsnAsnAsnSerThrLeuHisMetLeu----- 103  
QY 370 CTTCTCATGATTTTACCTTCAAGAGAAATCAACTTCATTC-----TCAACA 420  
Db 104 -----LeuProAsnHisGlnGlyPheAlaPheThrAspGluAsnThr 118  
QY 421 ATGCTTCCA-----TGGAATACCATCATGATCCTCTCTACAA 459  
Db 119 MetGlnProGlnGlnGlnHisPheThrTyrProSerSerSerAspHisGln 138  
QY 460 -----ATGGGTGGCTTTGATATTTCATTTCTATGCTGACTACAAATACTTA 507  
Db 139 AsnArgAspMetIleGlyThrValHisValGluGlyLysGlyLeuSerLeu 158  
QY 508 TCATCTTCTCCACGGTCTATCGATGTTCAAGATAACCGCAATGTTGA-----GTT 557  
Db 159 SerSerSerLeuAlaAlaAlaLysAlaGluGluTyrArgSerIleTyrCysAlaAlaVal 178  
QY 558 CATGGCTCTCTCTCATCT 608  
Db 179 AspGlyThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 198  
QY 609 ACATCATGATGATTCCTCAACAA----- 632  
Db 199 AsnLeu-----LeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 216  
QY 633 CATGTGGGGTTTTGAAGCAAAATAGTAGTTTCAGGCATTTTCAGGTGTAGTGTCTCAAG 692  
Db 217 His-PheGlySerSerSerSerProMetAlaAlaSerSerSerIleGlyGly----- 234  
QY 593 TGAACCAATGATGCTCTACATTCGATGAGAGAGATTTCCCGTTCTCTATTTTCAATAAAG 752

Db 235 -----ileTyThrLeuAr 239  
 QY 753 AAACAATGAGCTTTCATTGAGCTTGTGCATCAGATGTTCTTCATGAAATCGCGAGATAAG 812  
 Db 239 GAsn----- 240  
 QY 813 TCTTTGTGCAGCTACAAGATTAGCCTCAGACGAAGCTTCTTCAGCAGCAAAAGACATTC 872  
 Db 240 ----- 240  
 QY 873 TAATAACGTTGTACTCAGAGTTCTCTCAACTATATTTGGCTCAAAATACCTTCACCTC 932  
 Db 241 -----SerLysTyThrLysPr 246  
 QY 933 TGTTCAGAAATACATCTCATTTCCGCCGATCTCGCTCGATTATTCATCTCGAGGAAC 992  
 Db 246 oAlaGlnGluLeuGluGluPheCysSerValGlyArgGlyHisPheLysLysAsnLy 266  
 QY 993 CGAGTCAGAGCTGCTAGTTCAGCCTTTACTTCAGCTTTTCAGATATATACTAGTTCT 1052  
 Db 266 sLeuSerArgAsnAsnSerAsn-----ProAsnThrThrGlyGly 280  
 QY 1053 TCATGCTGATTCTAATAACTCGAGCGGGTTTCGGA-----TCTAC 1094  
 Db 280 yGlyGlyGlySerSerSerSerAlaGlyThrAlaAsnAspSerProProLeuSerPr 300  
 QY 1095 ATTCAAAGGAGCATTTAGAGCAAGAAACCCATCTCTTGGATCTCTTCAAATGGT 1154  
 Db 300 oAlaAspArgIleGluHisGlnArgGlyValLysLeuLeuSerMetLeuGluGluVa 320  
 QY 1155 GAGTATCATATAGTCTGCTGATGATGAGATCATACGCTTATATCAGCTTCCATGCG 1214  
 Db 320 lAspArgArgTyAsnHisTyrCysGlnGlnMetGlnMetValValAsnSerPheAspGl 340  
 QY 1215 TCAACCGAGTTAGATCCACAGTTA---CACACCGGTTTGCCCTCCAAACCGTTTCCTT 1271  
 Db 340 nValMetGlyTyArgAlaAlaValProTyThrThrLeuAlaGlnLysAlaMetSerAr 360  
 QY 1272 CTTATACAGACCTGAGAGAGATCTGCGAAGATATC----- 1314  
 Db 360 gHisPheArgCysLeuLysAspAlaValAlaValGlnLeuLysArgSerCysGluLeuLe 380  
 QY 1315 -----TCTATGCGATCTGTTATTCAGAGAGCGC----- 1341  
 Db 380 uGlyAspLysGluAlaAlaGlyAlaAlaSerSerGlyLeuThrLysGlyGluThrProAr 400  
 QY 1342 -----AAGACACACTCAAGAACCTCTATGTTCCACAGCATTCCTCTCT 1388  
 Db 400 gLeuArgLeuLeuGluGlnSerLeuArgGlnGlnArgAlaPheHisMetGlyMetMe 420  
 QY 1389 TCAGCAGCTGAAACGAAACCACTCAGATTGGAGACTCAACAGAGCTTGGCTCGAGAA 1448  
 Db 420 tGlnGln-----GluAlaTrpArgProGlnArgGlyLeuProGluAr 434  
 QY 1449 ATCTGTTCCGTTCTACGGAATGATGTTCCAAATCTCTTCCCTTCCCTTCCCTCCGAAAGA 1508  
 Db 434 gSerValAsnIleLeuArgAlaTrpLeuPheGluHisPheLeuAsnProTyProSerAs 454  
 QY 1509 TTCGGAGAACATCTTCTAGCTATAGAGTGGCTTGCAGCAAGCTCAGGTATCAAACTG 1568  
 Db 454 pAlaAspLysHisLeuLeuAlaArgGlnThrGlyLeuSerArgAsnGlnValSerAsnTr 474  
 QY 1569 GTTTATAATCGCGGGTTAGCTATGGAAGCCGATGATAGAGAGATGTAT----- 1620  
 Db 474 pPheIleAsnAlaArgValArgLeuTrpLysProMetValGluGluMetTyArgGlnGln 494  
 QY 1621 -----GCGGAATGACAGAGGAGCTCAATAACAGTCACATTC 1661  
 Db 494 uAlaLysGluArgGluGluAlaGlu--GluGluAsnGlnGlnGlnGlnArgArg 513  
 QY 1662 ACCCAACGCAACCTCTCTGATGCGCAAAATCTGTTATGATGAGCCCAAGCAAT 1715  
 Db 514 GlnGlnGlnThrAsnAsnAsnAspThrLys-----ProAsnAsn 526

## RESULT 11

D84679

probable homeodomain transcription factor [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C/Accession: D84679

R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; A

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.B.; Umayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A/Reference number: A84420; MUID:20083487; PMID:10617197

A/Accession: D84679

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-584 &lt;STO&gt;

A/Cross-references: GB:AE002093; NID:g4510417; PIDN:AAD21503.1; GSPDB:GN00139

C/Genetics:

A/Gene: At2g27990

A/Map position: 2

Alignment Scores:

Pred. No.: 1,08e-23 Length: 584

Score: 401.00 Matches: 139

Percent Similarity: 42.42% Conservative: 82

Best Local Similarity: 26.68% Mismatches: 172

Query Match: 12.21% Indels: 128

DB: 2 Gaps: 19

US-09-423-575-1 (1-1886) x D84679 (1-584)

QY 250 AATCTTCCCAAGGTTCTATCAGAGAGAGATTAAGTTTCATAGAACCAATGGAC 309

Db 70 AsnMetProGln-----SerIleHisArgAspProGlnGly 81

QY 310 AACACCAAC----- 318

Db 82 ProSerAsnTrpArgIleSerAspLeuSerGlnProSerThrValAsnHisGlyTyArg 101

QY 319 -----AACACCAACACTTTTACTCTCTCGATAATGTCATCAACCA 363

Db 102 GlnAlaGlyIleArgProAsnAsnValAlaAspLeuLeuSerAspHisPheSerArg 121

QY 364 AATCTCTCTCTCATGATTTTATACCTTCAAGAGAGATTCAACTTCATCTCAACAATG 423

Db 122 AsnGlnIleLeuAspArgProLeuTyValGlyArgAspSerIleProGlnSerMet 141

QY 424 CTTCCATGGAATACCATCAGATCA-----GATCCTCTCAAAATGGTGGCTTTGATATTTTC 480

Db 142 IleArgArgSerGluValSerCysLeuAspAspAsnGlnLysGlyCysValThrValAla 161

QY 481 AATTCTATGCTGACTAACAAATACATTATCTTCTCCACGGTCTATCGATGTTCAAGAT 540

Db 162 CysSerGlyThrGlyAsnGlnIleLeuArgSerSerTyArgPdnGlnGlySerSerGly 181

QY 541 AACCGCAATGTTGAGTTCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600

Db 182 SerTyArgGlyGluPheSerPheLeuPro-----SerLeuGlu 194

QY 601 CATTTAAGACACATCATGATCTCTCAAAACACATCTGGGTTTGAAGCAAAATAGTAG 660

Db 195 AsnGlnSerValAlaHisAsnAlaSerAsn-----TrpAsnHis----- 207

QY 661 TTTCCAGGCATTTTCAGGTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 714

Db 208 -----GlyProValAsnValThrAlaThrSerHisThrAsn 219

QY 715 GGTGAAGAGATTTCCGTTTCTTAATTCGAATTAAGAGAACAAATGAGCTTTTCATTGAGT 774

Db 220 SerLysLysGlyPhePro-----LeuSer 227

QY 775 CTTGTCATCAGATGTTTCTGATGATGCTCGAGATAAGTCTTTGTGTCAGCTCAAGATTA 834

```

Db      228 LeuLeuSerAspIlePro----- 233
QY      835 GCCTCAGACGAGCTTCTTGCACGACAAAGACATTTCTAATAACGTTGTTACTCAA---- 891
Db      234 -----ProSerArgAspValGlyAsnAlaValLeuSerThr 246
QY      892 -----GGTTTCTCAACTTATATTTGGCTCA 918
Db      247 MetAsnIleHisGlyProLeuGlyProPheThrGlyTrpAlaSerIleLeuLysSerSer 266
QY      919 AATACTCTCACTCTCTTCAAGAAATACATCTATTCATTTCCGCGCATACCTCGCTCCGATTAT 978
Db      267 ArgPheLeuGluProAlaGlnLysMetLeuGluGluPheCys-----IleSerThr 283
QY      979 TCATCTCGAGGACCGAGTCGAGAGCTGTAGTTCAGCCTTTACTTACCTCAGCTTTTGAGAAAT 1038
Db      284 AlaSerLys-----IleIleSerArgSerGluSer 293
QY      1039 ATAACCTGAGTTCTTGTGATGTTGATTCATAAATCGAGCGGGTTTCGATCTACATTT 1098
Db      294 ThrSerMetGluAspAspAspAspAspAsnLeuSerGlyPheSerSerSerSer 313
QY      1099 CAAGAGAGACATTAGAA-----GCAAGAGAAACCCATCTCTCTTGATCTTCTT 1146
Db      314 Glu-----ProLeuGluProLysAsnArgLeuLysLysAlaLysLeuLeuPheLeuGln 331
QY      1147 CAAATGGTGGATGATCGATATAGTCATTCGCTAGATGAGATTCATACCGTTATATACGCG 1206
Db      332 GluGluValCysLysTrpTyrlsLeuTyrlsAsnHisGlnLeuGlnThrValMetSerSer 351
QY      1207 TTCCATGCTGCAACCGAGTTAGATCCACAGTTA---CACACCGGTTTGGCCCTCCAAACC 1263
Db      352 PheAsnThrValAlaGlyLeuAsnThrAlaThrProTyrlsIleSerLeuAlaLeuLysArg 371
QY      1264 GTTTCCTCTTATACAGAACCTGAGAGAGAACTCAGAGAAACCTCTATGTTCCACGAGCATTCG 1323
Db      372 ThrSerArgSerPheLysAlaLeuArgThrAlaIleAlaGluHisValLysGlnIleSer 391
QY      1324 TCTGTATTGGAGAGAGCAAGACAGACTCAAGAAACCTCTATGTTCCACGAGCATTCG 1383
Db      392 SerHisSerSerAsnGlyAsnAsnAsnArgPheGlnLys-----ArgGlnArgSer 409
QY      1384 CTTCTTCAGCAG-----CTGAACGAAAGAACCATCATGTTGGAGACCTCAACGA 1434
Db      410 LeuIleGlyAsnAsnValGlyPheGluSerGlnGlnGlnHisIleTrpArgProGlnArg 429
QY      1435 GGTTCCTGCTGAGAAATCTGTTTCGTTCTACGGAATTCGATGTTCCAAACTTCCTTCAC 1494
Db      430 GlyLeuProGluArgAlaValAlaValLeuArgAlaTrpLeuPheAspHisPheLeuHis 449
QY      1495 CTTTACCGAAGATTCGAGNAACATCTTCTAGCTATACGAAGTGGCTTCAACAAGAGT 1554
Db      450 ProTyrlsProThrAspSerAspLysGlnMetLeuAlaThrGlnThrGlyLeuSerArgAsn 469
QY      1555 CAGGTATCAACTGTTTATAATCGCGGTTAGGCTATGAGCCGATGATGAGCCGATGAGAGAG 1614
Db      470 GlnValSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetValGluGlu 489
QY      1615 ATGTATCGGAAATGAACAG-----AGGAAGCTCAATAACAGTCAAC---ATTCAACCC 1665
Db      490 IleHisThrLeuGluThrLysAlaIleLysAsnAlaAspThrSerHisAsnIleGluPro 509
QY      1666 AAC 1668
Db      510 Ser 510

```

RESULT 12

T48224

probable homeodomain protein - Arabidopsis thaliana

N/Alternate names: protein T7H20.80

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C/Accession: T48224  
 R/Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; New  
 submitted to the Protein Sequence Database, March 2000  
 A/Reference number: 224488  
 A/Accession: T48224  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-575 <BEV>  
 A/Cross-references: EMBL:AL162508  
 A/Experimental source: cultivar Columbia; BAC clone T7H20  
 C/Genetics:  
 A/Map position: 5  
 A/Introns: 240/3; 374/2; 394/3  
 A/Note: T7H20.80

## Alignment Scores:

Pred. No.: 4.95e-22 Length: 575  
 Score: 380.00 Matches: 126  
 Percent Similarity: 40.72% Conservative: 67  
 Best Local Similarity: 26.58% Mismatches: 165  
 Query Match: 11.57% Indels: 116  
 DB: 2 Gaps: 14

US-09-423-575-1 (1-1886) x T48224 (1-575)

```

QY      532 GTTCAAGATAACCGCAATGTTGAGTTTCATGCTCCT----- 567
Db      11 LeuGlnGlnSerArgArgAspLysLeuArgIleProSerLeuAspSerHisPheHisPhe 30
QY      568 ---CCTCCTCATCCTCCT--- 597
Db      31 HisProProProProSerSerGlyGlyGlyGlyValPheProLeuAlaAspSer 50
QY      598 GATCATTTAAGACACATGATGATTCCTCAACAACATGTTGGGTTTGAAGCAATAGT 657
Db      51 AspPheLeuAlaGlyGlyPheHisSerAsnAsnAsnHisIleSerAsnPro 70
QY      658 GAGTTTCAGGCATTTTCAGGTGTAGTTGTT---CCAGTGAACCAATCATG 705
Db      71 SerTyrlsAsnPheMetGlyPheLeuGlyGlyProSerSerSerSerThrAlaVal 90
QY      706 TCTACATTCGTTGAAGAGATTTCCTCGGTTTCTAATTCGAATAAAGAAACAATGAGCTT 765
Db      91 AlaValAlaGlyAspHisSerPhe-----AsnAlaGlyLeu 102
QY      766 TCATTGAGTCTTCATCAGATGTTTCTGATGAATGTCGAGAGATAAGTCTTTGTCAGCT 825
Db      103 SerSerGlyAspValLeuValPheLysProGluProLeuSerLeuSerSerHis 122
QY      826 ACAAGATTACCTCAGAG-----CAAGCTTCTTTCAGCAGC 861
Db      123 ProArgLeuAlaTyrlsAspLeuValValProGlyValValAsnSerGlyPheCysArgSer 142
QY      862 AAAGACATTTCTAATAACGTTGTACT----- 888
Db      143 AlaGlyAlaAlaAlaAlaAlaValThrIleAlaSerArgSerGlyProLeuGly 162
QY      889 -----CAAGGTTTCTCTCAACTTATATTTGGCTCAAAATACCTTCACTCTGTCAAGAA 942
Db      163 ProPheThrGlyTyrlsAlaSerIleLeuLysGlySerArgPheLeuLysProAlaGlnMet 182
QY      943 ATACTATCTCAATTCCCGCATACTCGCTCGAT---TATTCACTCTCGAGAACCGAGTCA 999
Db      183 LeuLeuAspGluPheCysAsnValGlyArgGlyIleTyrlsAspLysValIleAspAsp 202
QY      1000 GGAGCTGTAGTTTCAGCCTTACTTCCAGCTTTTGTAGAAATATACTGAGTTTCTTGATGT 1059
Db      203 AspAspSerSerLeuLeuPheAspProThrValGluAsnLeuCysGlyValSerAspGly 222
QY      1060 GATCTTAATAACTCGGAGCGGTTTCGGATCTACATTTCAAAGGAGAGACATTAGAGCA 1119
Db      223 GlyGlyGlyAspAsn-----GlyLys 229

```





```

Db 129 -----MetAlaHisProHisTyrHisArgLeuLeuAlaAlaTyrValAsnCysGln 145
QY 763 CTTTCATTGAGCTTCATGATGTTTCTGATGAATCTCGGAGATAAGTCTTTGTGCA 822
Db 146 ---LysValGlyAlaProGluValValAlaArgLeuGluGluAlaCysAlaSerAla 164
QY 823 GCTCAAGATTAGCTCAGCAAGCT-----TCTTCAGCAGCAAGACATTCT 873
Db 165 AlaThrMetAlaGlyGlyAspAlaAlaGlySerCysIleGlyGluAspProAla 184
QY 874 AATAAGTTGTTACTCAAGTTTCTCAACTATATTGCTCAAAATACCTTCACTCT 933
Db 185 LeuAspGlnPheMetGluAlaTyrCysGluMetLeu-----ThrLysTyrGluGlnGlu 202
QY 934 GTTCAAGAAATACTATCTCATTTCCCGCATACTCGCTCGATTATTATCTCGAGGAACC 993
Db 203 LeuSerLysProLeuLys----- 208
QY 994 GAGTCAGGAGCTGTAGTTACGCTTTACTTACGTTTGTGAGATATAACTGAGTTCTT 1053
Db 209 GluAlaMetLeuPheLeuGlnArgIleGluCysGlnPheLysAsnLeuThr---IleSer 227
QY 1054 GATGTGATTCTTAATACTCGGAGCGGTTTCGGATCTACATTTCAAAGGAGAGCATTA 1113
Db 228 SerSerAspPheAlaSerAsnGluGly----- 237
QY 1114 GAAGCAAGAAACCCATCTCTTGGATCTTCTCAATGGTGGATGATCGATATAGTCAT 1173
Db 238 -----AspArgAsnGlySer 242
QY 1174 TCGGTAGATGAGATT-----CATACGGTTATATCAGCGTTTCCATGCTGCAACCGAGTTA 1227
Db 243 SerGluGluAspValAspLeuHisAsnMetIle----- 253
QY 1228 GATCCACAGTTACACACCGGTTGGCCCTCCAAACGGTTTCTTCTTATACAGAACCTG 1287
Db 254 AspProGlnAlaGluAspArg-----AspLeu 262
QY 1288 AGAGAGAGATCTGCAAGAGATAATCTCT---ATGGGATCTGTATTGGAGAGAGGCAAA 1344
Db 263 LysGlyGlnLeuLeuArgLysTyrSerGlyTyrLeuGlySer----- 276
QY 1345 GACAAGACTCAAGAAACCTCTATGTTCCACAGCATTTGCCCTTCTCAGCAGCTGAACGA 1404
Db 277 -----LeuLysGlnGluPheMetLys 283
QY 1405 AGAACCATCAGATTGGAGACCTCAACGAGGTTTGCTTGAGAAATCTGTTTCGGTTCTA 1464
Db 284 Lys-----ArgLysLysGlyLysLeuProLysGluAlaArgGlnGlnLeu 298
QY 1465 CGGAATTGGATTTCAAAACCTTCTTACCTTACCCGAAAGATTGGAGAAACATCTT 1524
Db 299 LeuGluTyrTrpAsnArgHisTyrLysTrpProTyrProSerGluSerGlnLysLeuAla 318
QY 1525 CTAGCTATACGAGTGGCTTGACAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1584
Db 319 LeuAlaGluSerThrGlyLeuAspGlnLysGlnIleAsnAsnTrpPheIleAsnGlnArg 338
QY 1585 GTTAGGCTATGGAAGCCGATGATAGAGAGATGATGCGGAAATGAAC 1632
Db 339 LysArgHisTrpLysProSerGluAspMetGlnPheValMetAsp 354

```



OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 2, 2004, 09:48:42 ; Search time 22 Seconds  
(without alignments)  
8927.665 Million cell updates/sec

Title: US-09-423-575-1  
Perfect score: 3284  
Sequence: 1 atttggtataaatgttgc.....ttaaaaaaaaaaaaaaaaaaaaaa 1886

Scoring table: BLOSSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-Q/cgn2\_1/USPTO\_spool/US09423575/runat\_01092004\_161106\_18556/app\_query.fasta\_1.2055  
-DB=SwissProt\_42 -QFMT=fastcan -SUFFIX=rep -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09423575 @CNG 1.1 26 @runat\_01092004\_161106\_18556 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -Ygapop=10 -Ygapext=10 -DELOP=6 -DELEXT=7

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2484	75.6	473	ATH1_ARATH	P48731 arabidopsis
2	208	6.3	435	PNX1_MOUSE	O70477 mus musculus
3	202	6.2	435	PNX1_HUMAN	P55347 homo sapien
4	196.5	6.0	390	ME11_MOUSE	Q60954 mus musculus
5	196.5	6.0	477	ME12_HUMAN	O14770 homo sapien
6	196.5	6.0	477	ME12_MOUSE	P97367 mus musculus
7	195.5	6.0	390	ME11_HUMAN	O00470 homo sapien
8	194	5.9	378	ME13_MOUSE	P97368 mus musculus
9	189.5	5.8	390	ME11_XENLA	P79937 xenopus lae
10	186.5	5.7	460	PNX2_HUMAN	P99687 homo sapien
11	182	5.5	382	ME13_HUMAN	Q96kn3 homo sapien
12	177	5.4	379	HMB1_SOYBN	P46608 glycine max
13	175.5	5.3	355	HKJ6_LYCES	O22289 lycopersico
14	175.5	5.3	431	HKJ3_ARATH	P48000 arabidopsis
15	169.5	5.2	364	KNX3_HORVU	Q43484 hordeum vul
16	169	5.1	291	SIX2_HUMAN	Q9np88 homo sapien
17	167	5.1	398	HKJ1_MALDO	O04134 malus domes
18	166.5	5.1	359	HKJ1_MAIZE	P24345 zea mays (m

19	166.5	5.1	659	1	HMN1_DROME	P22807 drosophila
20	161.5	4.9	361	1	OSH1_ORYSA	P46609 oryza sativ
21	161.5	4.9	398	1	HKJ1_ARATH	P46639 arabidopsis
22	158.5	4.8	419	1	HKJ4_ARATH	P48001 arabidopsis
23	158	4.8	296	1	SIX2_MOUSE	Q62232 mus musculu
24	157.5	4.8	269	1	AK3_CHICK	Q90655 gallus gall
25	157.5	4.8	276	1	YGF6_YEAST	P35147 saccharomyc
26	154	4.7	1314	1	SMW1_YEAST	P09547 saccharomyc
27	153	4.7	294	1	HD1_BRANA	P46606 brassica na
28	152.5	4.6	306	1	CUP9_YEAST	P41817 saccharomyc
29	152.5	4.6	397	1	HKJ2_MALDO	O41135 malus domes
30	150	4.6	272	1	TGIF_HUMAN	Q15583 homo sapien
31	149.5	4.6	88	1	HKJ3_MAIZE	P56661 zea mays (m
32	149.5	4.6	272	1	TGIF_MOUSE	P70284 mus musculu
33	149.5	4.6	355	1	HKJ1_LYCES	Q41330 lycopersico
34	146	4.4	273	1	SIX1_MOUSE	Q62231 mus musculu
35	146	4.4	284	1	SIX1_HUMAN	Q15475 homo sapien
36	145	4.4	383	1	HKJ5_ARATH	P48002 arabidopsis
37	144	4.4	430	1	PEX1_HUMAN	P40424 homo sapien
38	144	4.4	430	1	PEX1_MOUSE	P41778 mus musculu
39	143.5	4.4	427	1	HKJ3_MALDO	O04136 malus domes
40	143	4.4	329	1	HM40_CAEEL	Q19503 caenorhabdi
41	142.5	4.3	351	1	RSH1_MAIZE	Q41853 zea mays (m
42	142	4.3	382	1	STM_ARATH	Q38874 arabidopsis
43	141.5	4.3	416	1	SO_DROME	O27350 drosophila
44	140.5	4.3	330	1	PEX4_HUMAN	Q9hyul homo sapien
45	140.5	4.3	426	1	HKJ6_LYCES	O23300 lycopersico

#### ALIGNMENTS

##### RESULT 1

ID	ATH1_ARATH	STANDARD;	PRT;	473 AA.
AC	P48731;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Homeobox protein ATH1			
GN	ATH1 OR AT4G32980 OR F26P21.100.			
OS	Arabidopsis thaliana (mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia; TISSUE=Seedling;			
RX	MEDLINE=95210927; PubMed=7696878;			
RA	Quaedvlieg N., Dockx J., Rook F., Weisbeek P.J., Smeekens S.;			
RT	"The homeobox gene ATH1 of Arabidopsis is derepressed in the			
RL	photomorphogenic mutants cop1 and det1.";			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia;			
RX	MEDLINE=20083488; PubMed=10617198;			
RA	Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,			
RA	Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,			
RA	Harris B., Ansoerger W., Brandt P., Grivell L.A., Bieger M.,			
RA	Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,			
RA	Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidheini T.,			
RA	Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,			
RA	Vos P., Hobeisel J., Zimmermann W., Wedler H., Ridley P.,			
RA	Langham S.-A., McCullagh B., Bilham L., Robben J.,			
RA	Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche P.,			
RA	Braeken M., Weljens I., Voet M., Bastiaens I., Aert R., Defoor E.,			
RA	Weitzenecker T., Bothe G., Ransperger U., Hilbert H., Braun M.,			
RA	Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,			
RA	Moolenaar P., Klein lankhorst R., Rose M., Hauf J., Koetter P.,			
RA	Berneiser S., Hempel S., Feldpausch M., Lambeth S., Van den Daele H.,			
RA	De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,			
RA	Van Montagu M., Rogers J., Cronin J., Quail M.A., Bray-Allen S.,			







```
QY 1180 GATCAGATTTCATACGGTTATATCAGCGTTCCATGCTGCAACCGAGTTAGATCCACAG-----1236
Db 173 GlnGlnleGlnSerAlaIleThrGly-:::|||||:::
QY 1237 -----TTACACACCGCGTTTGCCTCCAAACCGTTTCC-----1269
Db 188 IleValValProAlaSerAlaLeuGlnGlnGlnGlyAsnValAlaMetAlaThrValAlaGly 207
QY 1270 ----TTCCTTATACAGAACCTG-----AGAGAGAGAACTGCGAAGAGATA 1311
Db 208 GlyThrValTyrGlnProValThrValThrProGlnGlnGlnValValThrGlnThr 227
QY 1312 ATCTCTATGGGATCTGTATTGGAGAGAGGCAAA-----CACAGACTCAAGAA 1359
Db 228 LeuSerProGlyThrIleArgIleGlnAsnSerGlnLeuGlnLeuGlnLeuGlnAsp 247
QY 1360 ACCTCTATGTTCCACGAGATTCCTCTTCACGACTGAAACGAAAGAACCATCATCAT 1419
Db 248 LeuSerIleLeuHisGln-----AspAspGlySerSerLysAsn-----260
QY 1420 TGGAGACCTCAACGAGGT--TTGCCTGAGAAATCTGTTCCGTTTACGGAATTCGATG 1476
Db 261 -----LysArgGlyValLeuProLysHsAlaThrAsnValMetArgSerTrpLeu 277
QY 1477 TTCACAAATCTCTCACCTTACCCGAAAGATTTCGAGAAACATCTTCTAGCTATACGA 1536
Db 278 PheGlnHisIleGlyHisProTyrProThrGluAspGluLysGlnIleAlaGln 297
QY 1537 AGTGGCTTACAGAGAGTCAAGTATCAACTGTTTATAAATCGCGGTTAGGCTATGG 1596
Db 298 ThrAsnLeuThrLeuLeuGlnValAsnAsnTrpPheIleAsnAlaArgArgIleLeu 317
QY 1597 AAGCCGATGATGACAGAGATGTATGCGGAAATGAACAAGAGAGAGTCAATAACAGTCTAC 1656
Db 318 GlnProMetLeuAspSerSerCysSerGluThrProLysThrLysLysThrAlaGln 337
QY 1657 ATCAACCC-----AAGGACCAACTCTTCCGAATG 1686
Db 338 AsnArgProValGlnArgPheTrpProAspSerIleAlaSerGlyValAlaGlnProPro 357
QY 1687 CCAAAATCTGTATGATGAGCCAA 1710
Db 358 ProSerGluLeuThrMetSerGlu 365

RESULT 4
MEI1_MOUSE
ID MEI1_MOUSE STANDARD; PRT; 390 AA.
AC Q60954; Q60955;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Homeobox protein Meis1 (Myeloid ecotropic viral integration site-1).
GN MEI1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS MEIS1A AND MEIS1B).
RX MEDLINE=96009570; PubMed=7565694;
RA Moskow J.J., Bullrich F., Huebner K., Daar I.O., Buchberg A.M.;
RT "Meis1, a PBX1-related homeobox gene involved in myeloid leukemia in
RT BXH-2 mice.";
RL Mol. Cell. Biol. 15:5434-5443 (1995).
CC -!- FUNCTION: May function as a cofactor for HOXA7 and HOXA9 in the
CC induction of myeloid leukemias.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=alternative splicing; Named isoforms=2;
CC Comment=additional isoforms seem to exist;
CC Name=Meis1A;
CC IsoId=Q60954-1; Sequence=Displayed;
CC Name=Meis1B;
```

```
CC IsoId=Q60954-2; Sequence=VSP_002240;
CC -!- TISSUE SPECIFICITY: Expressed at high levels in the lung with
CC lower levels detected in the heart and brain.
CC -!- DEVELOPMENTAL STAGE: Expressed at high levels in all stages of
CC embryonic development analyzed (7 days to 17 days).
CC -!- INDUCTION: Expression is coactivated by retroviral integration in
CC BXH-2 murine myeloid leukemias.
CC -!- DISEASE: MEIS1 serves as a site of viral integration in 15% of the
CC tumors arising in BXH-2 mice that develop myeloid leukemia as a
CC result of the expression of an ecotropic murine leukemia virus.
CC -!- SIMILARITY: Belongs to the TALE/MEIS homeobox family.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U33629; AAA85508.1; -.
CC EMBL; U33630; AAA85509.1; -.
CC HSSP; P40424; 1B72.
CC TRANSFAC; T03388; -.
CC TRANSFAC; T03389; -.
CC MGD; XGI:104717; Meis1.
CC GO; GO:0005634; C:nucleus; IDA.
CC GO; GO:0005667; C:transcription factor complex; IDA.
CC GO; GO:0003677; F:DNA binding; IDA.
CC GO; GO:0005515; F:protein binding; IDA.
CC GO; GO:0003705; F:RNA polymerase II transcription factor acti. ; IDA.
CC GO; GO:0006357; P:regulation of transcription from Pol II pro. ; IC.
CC InterPro; IPR001356; Homeobox.
CC Pfam; PF00046; Homeobox; 1.
CC ProDom; PD00010; Homeobox; 1.
CC SMART; SM00389; HOX; 1.
CC PROSITE; PS00027; HOMEBOX 1; FALSE_NEG.
CC PROSITE; PS00071; HOMEBOX 2; 1.
CC DNA-binding; Nuclear protein; Homeobox; Proto-oncogene;
CC KW Alternative splicing.
CC FT DOMAIN 194 240 SER/THR-RICH.
CC FT DOMAIN 242 269 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 262 269 POLY-ASP.
CC FT DNA BIND 272 334 HOMEBOX (TALE-TYPE).
CC FT VARSPIC 373 390 PMSGGMNMGEGOWHYM -> LOSMPEYVARGPVGVM
CC QPSPYQACQMPHPAQLRHGPPHTYIPGHHPHAPVMMHG
CC isoform Meis1B).
CC /FTId=VSP_002240.
CC SQ SEQUENCE 390 AA; 43002 MW; E0C32B5CE25E1E2C CRC64;

Alignment Scores:
Pred. No.: 1.14e-07 Length: 390
Score: 196.50 Matches: 97
Percent Similarity: 37.40% Conservative: 50
Best Local Similarity: 24.68% Mismatches: 144
Query Match: 5.98% Indels: 102
DB: 1 Gaps: 14

US-09-423-575-1 (1-1886) x MEI1_MOUSE (1-390)
QY 574 CATCCTCTCCACTT-----CATCCTTGGATCATTTAAGACACTATGATGATTCCTCA 627
Db 41 HisGlyProLeuHisSerHisGlnTyrProHisThrAlaHisThrAsnAlaMetAla 60
QY 628 AACACATGTGGGTTTGAAGCAATAGTACGTTTCAGGCATTTTCAGGTGTAGTTGGT 687
Db 61 ProSerMet---GlySerSerValAsnAspAlaLeuLysArgAspLysAspAlaIle--- 78
QY 688 CCAAGTGACCAATGATGTTCTACATTCGGTGAAGAAGATTTCCTGTTTCTAATTTCGAAT 747
Db 79 -----TyrGlyHisProLeuPheProLeuLeuAlaLeuLe 90
```

QY	748	AAAAGAAACATGACGCTTTTCATGAGTCTTGATCAGATGTT-----TCTGAT	795
Db	91	PheGluLysCysGluLeuAlaThrCysThrProArgGluProGlyValAlaGlyGlyAsp	110
QY	796	GAATGCTCGACATAAGTCTT-----TGTGACGCTCAACAGATTAGCTCTCA	840
Db	111	ValCysSerSerGluSerPheAsnGluAspIleAlaValPheAlaLysGluIleAla	130
QY	841	GAGCAAGCTTCTTGACGACGACAAACATCTTCTAATACGTTGTACTCAAGTTCTCT	900
Db	131	GluLysProLeuPheSerSerAsnProGluLeuAspAsnLeuMetIleGlnAlaIleGln	150
QY	901	CAACTATATTGGCTCAAAATACCTTCACTCTGTTCAAGAAATACTATCTCATTTTC	957
Db	151	ValLeuArgPheHisLeuLeuGluLysValHisGluLeuCysAspAsnPheCys	170
QY	958	-----GCCGCATACCTCGCTCGATTATTCATCTCGAGGAACC	993
Db	171	HisArgTyrIleSerCysLeuLysGlyLysMetProIleAspLeuValIleAspArg	190
QY	994	GAGTCAGGAGCTGCTAGT---TCAGCCTTACTTCCAGCTTTTGACAATATACTAGTTT	1050
Db	191	GluGlyGlySerLysSerAspSerGluAspValThrArgSerAlaAsnLeuThr	208
QY	1051	CTTGATGTTGATTTAATAAATCACTCGAGGCGGTTTCGGATCTACATTTCAAAGGAGACA	1110
Db	209	---AspGlnProSerTrpAsn-Arg-----Asp-HisAspAspThrAlaSerT	223
QY	1111	TTAGAGCAAGAAAGAACCCA-----TCTCTGGATCTCTCAATGGTGGATGAT	1161
Db	223	hrArgSerGlyGlyThrProGlyProSerSerGlyGlyHisThrSerHisSerGly-Asp	242
QY	1162	CGATATAGTCAITTCGTATAGATGATATACGGTTATATACGGTTCATCGCTCAACC	1221
Db	243	AsnSerSerGluGlnGlyAspGlyLeuAspAsnSerValAlaSerProSerThrGlyAsp	262
QY	1222	GAGTTAGATCCAGGTATACACACCCCGTTGCGCCCTCCCAACCGTTTCTCTTATACAAG	1281
Db	263	AspAspAspProAsp-----	267
QY	1282	AACCTGAGAGAGATCTGCAAGAAGATACTCTATGGATCTGTATTGGAGAGAGCC	1341
Db	267	-----	267
QY	1342	AAAGACAAGACTCAAGAAACCTCTATGTTCCACAGCATTGCTTCTTCAGCAGCTGAAA	1401
Db	268	LysAspLys-----	270
QY	1402	CGAAGAACCATCAGATTTGGAGACCTCAACGAGTTTG---CCTGAGAAATCTGTTCCG	1458
Db	271	---LysArgHisLys-----LysArgGlyIlePheProLysValAlaThrAsn	285
QY	1459	GTCTACGGAATGGATGTTTCCAAAACCTTCCTTACCCCTTACCCGAAAGATTCGGAGAA	1518
Db	286	IleMetArgAlaTrpLeuPheGlnHisLeuThrHisProTyrProSerGluGluGlnLys	305
QY	1519	CATCTTCTAGCTATAGAGTGGCTTGACAGAAGTCAGGTATCAACTGTTTATAAT	1578
Db	306	LysGlnLeuAlaGlnAspThrGlyLeuThrIleLeuGlnValAsnAsnTrpPheIleAsn	325
QY	1579	GCCTGGGTTAGGCTATGGAAGCCGATGATAGAGATGATGATCGGAAATGAACAGAGG	1638
Db	326	AlaArgArgArgIleValGlnProMetIleAspGlnSerAsnArgAlaValSerGlnGly	345
QY	1639	AACTCAATAAACAGTCACATTCACCCACGGA	1671
Db	346	ThrProTyrAsnProAspGlyGlnProMetGly	356
AC	MEIS2_HUMAN	STANDARD; PRT; 477 AA	
AC	O14770; Q96K14; Q96K15; Q9NRS1; Q9NRS2; Q9NRS3;		

RESULT 5  
MEIS2\_HUMAN  
ID MEIS2\_HUMAN  
AC O14770; Q96K14; Q96K15; Q9NRS1; Q9NRS2; Q9NRS3;

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL; AF178948; AAF81638.1; -  
DR EMBL; AF179896; AAF81639.1; -  
DR EMBL; AF179897; AAF81640.1; -  
DR EMBL; AF179898; AAF81641.1; -  
DR EMBL; AF179899; AAF81642.1; -  
DR EMBL; BC050431; AAF80431.1; -  
DR EMBL; AF017418; AAF80270.1; -  
DR HSP; P40424; I872.  
DR TRANSFAC; T04114; -  
DR TRANSFAC; T04115; -  
DR TRANSFAC; T04116; -  
DR TRANSFAC; T04117; -  
DR TRANSFAC; T04118; -  
DR Genew; HGNC:7001; MEI2.  
DR MM; 601740; -  
DR GO; GO:0003704; F:specific RNA polymerase II transcription fa. .; TAS.  
DR GO; GO:0003714; F:transcription co-repressor activity; TAS.  
DR GO; GO:0000122; P:negative regulation of transcription from P. .; TAS.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 1.  
DR ProDom; PD000010; Homeobox; 1.  
DR SMART; SM00389; Hox; 1.  
DR PROSITE; PS00027; HOMEBOX 1; FALSE\_NEG.  
DR PROSITE; PS00071; HOMEBOX 2; 1.  
KW Homeobox; DNA-binding; Nuclear protein; Alternative splicing.  
FT DOMAIN 195 244 SEP/THR-RICH  
FT DOMAIN 245 273 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 266 273 POLY-ASP.  
FT DNA\_BIND 276 338 HOMEBOX (TALE-TYPE).  
FT VARSPPLIC 346 352 Missing (in isoform Meis2B and isoform Meis2D).  
FT VARSPPLIC 301 302 /FTId=VSP 002242.  
FT VARSPPLIC 303 477 HP -> VY (in isoform Meis2E).  
FT VARSPPLIC 303 477 /FTId=VSP 002243.  
FT VARSPPLIC 384 401 Missing (in isoform Meis2B).  
FT VARSPPLIC 402 477 LQMPGDYVSGGPGMGMS -> PMSGVMNMGMGQWHYM (in isoform Meis2A and isoform Meis2B).  
FT VARSPPLIC 402 477 /FTId=VSP 002245.  
FT VARSPPLIC 402 477 Missing (in isoform Meis2A and isoform Meis2B).  
FT SEQUENCE 477 AA; 51790 MW; 94EBD0801A312B24 CRC64;  
Alignment Scores:  
Pred. No.: 1.18e-07 Length: 477  
Score: 196.50 Matches: 90  
Percent Similarity: 33.89% Conservative: 52  
Best Local Similarity: 21.48% Mismatches: 152  
Query Match: 5.98% Indels: 125  
DB: 1 Gaps: 12

US-09-423-575-1 (1-1886) x MEI2\_HUMAN (1-477)  
QY 460 ATGGGTGGTTCGATTTTCAATTTCTATGCTGCTAACAATACTATCATCTCTCCA 519  
Db 14 MetAspGlyValGlyValProAlaSerMetTyrGlyAspProHis-----AlaPro 30  
QY 520 CGGTCTATCGATGTTCAAGATAACCGCAATGTTGAGTTTCATCGCTCCTCCTCT 573  
Db 31 ArgProIle-----ProProValHis 37  
QY 574 -----CATCTCTCCATTCATCTTTGGATCATTTTAAGACACTATGATGATTC 624  
Db 38 HisLeuAsnHisGlyProProLeuHisAlaThrGlnHisTyrGlyAlaHisAlaProHis 57  
QY 625 TCAACACACATG-----TGGGGTTTGAAGCAATAGTAGTTCAGGCATTTTCA 675  
Db 58 ProAsnValMetProAlaSerMetGlySerAlaValAsnAlaLeuLeuArgAspLys 77

QY 676 GGTGTAGTGTGTCGAAGTGAACCAATGATGCTACATTCGGTGAAGAAGATTCCCGTTT 735  
Db 78 AspAlaIle-----TyrGlyHisProLeuPheProLeu 88  
QY 736 CTAATTTCCGAATAAAGAAACAATGAGCTTTTCATTGACT----- 774  
Db 89 LeuAlaLeuValPheGluLysCysGluLeuAlaThrCysThrProArgGluProGlyVal 108  
QY 775 CTTGCATCAGATGTT-----TCTGATGAATGCTCGGAGATAAGTCTTTGTCAGCTACA 828  
Db 109 AlaGlyGlyAspValCysSerSerAspSerPheAsnGluAspIleAlaValPheAlaLys 128  
QY 829 AGATTAGCTCAGACCAAGCTTCTGACGAGCAAGACATTTCTTAATAGCTGTGTTACT 888  
Db 129 GlnValArgAlaGluLysProLeuPheSerSerAsnProGluLeuAspAsnLeuMetIle 148  
QY 889 CAAGGTTTCTCTCAACTTATATTGGCTCAAAATACCTTCACTCTGTTCACAAATACTA 948  
Db 149 GlnAlaIleGlnValLeuArgPheHisLeuLeuGluLysValHisGluLeuCys 168  
QY 949 TCTCATTTCCGGCATCTCCGCTC----- 972  
Db 169 AspAsnPheCysHisArgTyrIleSerCysLeuLysGlyLysMetProIleAspLeuVal 188  
QY 973 -----GATTATTTCATCTCGAGAAACCGAGTCAGGAGCTGTAGTTACGCCCTT 1020  
Db 189 IleAspGluArgAspGlySerSerLysSerAspHisGluLeuSerGlySerSerThr 208  
QY 1021 ACTTCACCTTTTGAGATATAACTGATTTCTTGTGATGATTAATACTCGAGCGC 1080  
Db 209 AsnLeuAlaAspHisAsnProSerSerTrpArgAspHisAspAlaThrSerThrHis 228  
QY 1081 GGTTCGATCTACATTTCAAGGAGAGCATTAAGACCAAGAAACCCCTCTCTTGAT 1140  
Db 229 SerAlaGlyThrProGlyProSerSerGlyGlyHisAlaSerGlnSer----- 244  
QY 1141 CTTCTTCAATGTTGGATGATGATATAGTCAATTGCGTAGATGATGATTCATCGGTTATA 1200  
Db 245 -----GlyAspAsnSerSerGluGlnGlyAspGlyLeuAspAsnSerVal 259  
QY 1201 TCAGCGTTCATGCTGCAACCGAGTTAGATCCACAGTTACACCCCGTTCCTCCAA 1260  
Db 260 AlaserProGlyThrGlyAspAspAspProAsp----- 271  
QY 1261 ACCGTTTCTCTTATACAAGAACCTGAGAGAGAGAAATCTCAAGAGAGATAATCTCTATG 1320  
Db 271 ----- 271  
QY 1321 GGATCTGTATGGAGAGAGCAAGACAAGACTCAAGAAACCTCTATGTTCCACCAT 1380  
Db 272 -----LysAspLysLysArgGln----- 277  
QY 1381 TGCCTTCTTCAGCAGCTGAAACGAAAGAACCATCAGATTTGGAGACCTCAACGAGTTTG 1440  
Db 278 -----LysLysArgGlyLeu 282  
QY 1441 ---CCTGAGAAATCTGTTTCGGTTCTACGGAATGGATGTTCCAAACTTCCTTCACCT 1497  
Db 283 PheProLysValAlaThrAsnIleMetArgAlaThrLeuPheGlnHisLeuThrHisPro 302  
QY 1498 TACCCGAAAGATTCCGAGAGAACATCTTCTAGCTATACGAGTGGCTGACAAAGTCTAG 1557  
Db 303 TyrProSerGluGluGlnLysLysGlnLeuAlaGlnAspThrGlyLeuThrIleLeuGln 322  
QY 1558 GTATCAAACTGGTTTATAATGCCGGTTAGCGTATGGAAGCCGATGATGAGAAGAG 1614  
Db 323 ValAsnAsnTrpPheIleAsnAlaArgArgIleValGlnProMetIleAspGln 341  
RESULT 6  
MEI2\_MOUSE  
ID MEI2\_MOUSE STANDARD; PRT; 477 AA.  
AC P97367; O35676; P97403; P97404;  
DT 15-JUL-1998 (Rel. 36, Created)







Db	149	GlnAlaIleGlnValLeuArgPheHisLeuLeuGluLeuGluHisGluLeuCys	168
QY	949	TCTCATTTCCGCGCATCTCGTC-----	972
Db	169	AspAsnPheCysHisArgTyrIleSerCysLeuLysGlyLysMetProLeuAspLeuVal	188
QY	973	-----GATTATTCTATCTCGAGAACCGAGTCAGAGCTGCTAGTTCAGGCTTT	1020
Db	189	IleAspGluArgAspGlySerLysSerLysAspHisGluGluLeuSerGlySerSerThr	208
QY	1021	ACTTCAGTTTTCAGATATACTAGTTTCTTGATGGTGATCTTAATACTCGGAGGG	1080
Db	209	AsnLeuAlaAspHisAsnProSerSerThrArgAspHisAspAspAlaThrSerThrHis	228
QY	1081	GGTTTCGATCTACATTTCAAAGGAGGAGCATTAAGAAGCAAGAAACCCATCTCTTGGAT	1140
Db	229	SerAlaGlyThrProGlyProSerSerGlyGlyHisAlaSerGlnSer-----	244
QY	1141	CTTCTCAAATGGTGATGATGATATAGTCATTCGCTAGATGAGATCATCGGTTATA	1200
Db	245	-----GlyAspAsnSerSerGluGlnGlyAspGlyLeuAspAsnSerVal	259
QY	1201	TCAGCGTTCCATCTGCAACCGAGTTAGATCCACAGTTACACACCGGTTTCCCTCCAA	1260
Db	260	AlaSerProGlyThrGlyAspAspAspProAsp-----	271
QY	1261	ACGGTTTCTCTTATACAGAACCTGAGAGAGAGAAATCTGCAAGAGATAATCTCTATG	1320
Db	271	-----	271
QY	1321	GGATCTGTATTGGAGAGCGCAAGACAGACTCAAGAACCTCTATGTTCACACAGAT	1380
Db	272	-----LysAspLysLysArgGln-----	277
QY	1381	TGCTTCTTTCAGCAGCTGAAACGAAAGAACCATCAGATTGGAGACTCAAGAGGTTTG	1440
Db	278	-----LysLysArgGlyIle	282
QY	1441	---CCTGAGAAATCTGTTTCGGTTCACGAATGGATGTTTCAAAACTCTCTCACCT	1497
Db	283	PheProLysValAlaThrAsnIleMetArgAlaTrpLeuPheGlnHisLeuThrHisPro	302
QY	1498	TACCCGAAGATTCGGAGAAACATCTCTAGCTATACGAGTGGCTTGACAAAGATCAG	1557
Db	303	TyrProSerGluGluGlnLysLysGlnLeuAlaGlnAspThrGlyLeuThrIleLeuGln	322
QY	1558	GTATCAAACTGGTTTATAATCGCGGTTAGCTATGGAACCGGATGATAGAAG	1614
Db	323	ValAsnAsnTrpPheIleAsnAlaArgArgIleValGlnProMetIleAspGln	341
RESULT 7			
ID	MEIL_HUMAN	STANDARD; PRT; 390 AA.	
AC	O00470;		
DT	15-JUL-1998 (Rel. 36, Created)		
DT	15-JUL-1998 (Rel. 36, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Homeobox protein Meis1.		
GN	MEIS1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=9736938; PubMed=9226379;		
RA	Smith J.E. Jr., Bollekens J.A., Inghirami G., Takeshita K.;		
RT	"Cloning and mapping of the Meis1 gene, the human homolog of a murine		
RT	leukemogenic gene.";		
RL	Genomics 43:99-103(1997).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Uterus;		

RX	MEDLINE=2238257; PubMed=12477932;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.F., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong L.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,		
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,		
RA	Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Fahy J., Heltón E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,		
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,		
RA	Schneerch A., Schein J.E., Jones S.J.M., Marra M.A.;		
RT	"Generation and initial analysis of more than 15,000 full-length		
RT	human and mouse cDNA sequences.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		
CC	-!- SUBCELLULAR LOCATION: Nuclear (Probable).		
CC	-!- TISSUE SPECIFICITY: Expressed at low level in normal		
CC	immunohematopoietic tissues, including the fetal liver. Expressed		
CC	in a subset of myeloid leukemia cell lines, with the highest		
CC	expression seen in those with a megakaryocytic-erythroid phenotype.		
CC	Also expressed at high levels in the cerebellum.		
CC	-!- SIMILARITY: Belongs to the TALE/MEIS homeobox family.		
CC	-!- SIMILARITY: Contains 1 homeobox domain.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; U085707; AAC51642.1; -.		
DR	EMBL; BC043503; AAH43503.1; -.		
DR	HSSP; P40424; 1B72.		
DR	TRANSFAC; T03419; -.		
DR	Genew; HGNC; 7000; MEIS1.		
DR	MIM; 601739; HOX; 1.		
DR	GO; GO:0003702; F:RNA polymerase II transcription factor acti. .; TAS.		
DR	InterPro; IPR001356; Homeobox.		
DR	Pfam; PF00046; homeobox; 1.		
DR	ProDom; PD000010; Homeobox; 1.		
DR	SMART; SM00389; HOX; 1.		
DR	PROSITE; PS00027; HOMEBOX_1; FALSE_NEG.		
DR	PROSITE; PS50071; HOMEBOX_2; 1.		
KW	DNA-binding; Nuclear protein; Homeobox; Proto-oncogene.		
FT	DOMAIN 194 240 SER/THR-RICH.		
FT	DOMAIN 242 269 ASP/GLU-RICH (ACIDIC).		
FT	DOMAIN 262 269 POLY-ASP.		
FT	DNA_BIND 272 334 HOMEBOX (TALE-TYPE).		
SQ	SEQUENCE 390 AA; 43016 MW; 3B6A0ACCF0C39121 CRC64;		
Alignment Scores:			
Pred. No.:	1.36e-07	Length:	390
Score:	195.50	Matches:	97
Percent Similarity:	37.40%	Conservative:	50
Best Local Similarity:	24.68%	Mismatches:	144
Query Match:	5.95%	Indels:	102
DB:	1	Gaps:	14
US-09-423-575-1 (1-1886) x MEIL_HUMAN (1-390)			
QY	574	CATCCTCTCCACCTT-----CATCCTTTGGATCATTTAAGACACTATGATTCCTCA	627
Db	41	HisGlyProLeuHisSerHisGlnTyrProHisThrAlaHisThrAsnAlaMetAla	60
QY	628	AACAACATGTGGGGTTTGTGAAGCAAAATAGTAGTTTCAGGATTTTCAGGTAGTTGGT	687

Db 61 ProSerMet---GlySerValAsnAspAlaLeuLysArgAspLysAspAlaLeu--- 78  
QY 688 CCAAGTAGCAACATGATGCTACATTCAGTGAAGAGATTTCCTGTTTCTGATTTTGAAT 747  
Db 79 -----TyrGlyHisProLeuPheProLeuLeuAlaLeuLeu 90  
QY 748 AAAAGAAACAATGAGCTTTCATGAGCTGTCAGATGTT-----TCTGAT 795  
Db 91 PheGluLysCysGluLeuAlaThrCysThrProArgGluProGlyValAlaGlyGlyAsp 110  
QY 796 GAATGCTCGGATAGCTTT-----TGTGAGCTACAAAGATTAGCTCA 840  
Db 111 ValCysSerSerGluSerPheAsnGluAspIleAlaValPheAlaGlyGluIleArgAla 130  
QY 841 GAGCAAGCTTCTTGACGAGCAAGACATTTCTAATAACGTTGTTACTCAAGGTTTCTCT 900  
Db 131 GluLysProLeuPheSerSerAsnProGluLeuAspAsnLeuMetIleGlnAlaIleGln 150  
QY 901 CAACTTATATTGGCTCAAAATACCTTCTACTCTGTTCAAGAAATCTATCTCATTTTC--- 957  
Db 151 ValLeuArgPheHisLeuLeuGluLeuGluLysValHisGluLeuCysAspAsnPheCys 170  
QY 958 -----GCCGATACCTCGCTCGATTTATCTATCTCGAGGAACC 993  
Db 171 HisArgTyrIleSerCysLeuLysGlyLysMetProIleAspLeuValIleAspArg 190  
QY 994 GAGTCAGGAGCTGCTAGT---TCAGCTTATCTTACGTTTGTGAGAAATATACTGATTT 1050  
Db 191 GluGlyGlySerLysSerAspSerGluAspIleThrArgSerAlaAsnLeuThr--- 208  
QY 1051 CTTGATGGTGATTAATAACTCGAGGGGGGTTTCCGATCTACATTTCAAAGGAGACA 1110  
Db 209 ---AspGlnProSerTrpAsn-Arg-----Asp-HisAspAspThrAlaSerT 223  
QY 1111 TTAGAAGCAAGAAACCA-----TCTTGGATCTTCTTCAAATGGTGATGAT 1161  
Db 223 hrArgSerGlyGlyThrProGlyProSerSerGlyGlyHisThrSerHisSerGly-Asp 242  
QY 1162 CGATATAGTCATTCCTAGATGATTCATACCGTTATATACGTTTCCATGCTCAACC 1221  
Db 243 AsnSerSerGluGlnGlyAspGlyLeuAspAsnSerValAlaSerProSerThrGlyAsp 262  
QY 1222 GAGTTAGATCCACAGTTACACACCCCGTTTGCCTCCAAACCGTTTCTCTTATATCAAG 1281  
Db 263 AspAspAspProAsp-----LysArgHisLys----- 267  
QY 1282 AACCTGAGAGAGAGAACTGCAAGAGAAATACTCTATGGATCTGTATTGGAGAGAGC 1341  
Db 267 ----- 267  
QY 1342 AAAGACAAGACTCAAGAAACCTTATGTTCACACAGCATTCGCTTCTTCCAGAGCTGAAA 1401  
Db 268 LysAspLys----- 270  
QY 1402 CGAAAGAACATCAGATTTGGAGACCTCAACAGGTTTG---CCTGAGAAATCTCTTTTCG 1459  
Db 271 ---LysArgHisLys-----LysArgGlyIlePheProLysValAlaThrAsn 285  
QY 1459 GTTCTACGGAATTTGGATTTTCCAAACTTCCCTTACCTTACCCTTACCGAAATCGGAGAAA 1518  
Db 286 IleMetArgAlaAlaTrpLeuPheGlnHisLeuThrHisProTyrProSerGluGluGlnLys 305  
QY 1519 CATCTTCTAGCTATACGAATGGCTGTGACAGAAGTCAGTATCAAACTGGTTTATAAT 1578  
Db 306 LysGlnLeuAlaGlnAspThrGlyLeuThrIleLeuGlnValAsnAsnTrpPheIleAsn 325  
QY 1579 GCGCGGTTAGGCTATGGAACCGGATGATAGAGAGATGTATCGGAATGAACAGAGG 1638  
Db 326 AlaArgArgArgIleValGlnProMetIleAspGlnSerAsnArgAlaValSerGlnGly 345  
QY 1639 AAGCTCAATAACAGTCACATTTCAACCAACGGA 1671

Db 346 ThrProTyrAsnProAspGlyGlnProMetGly 356  
RESULT 8  
MEI3 MOUSE  
ID MEI3 MOUSE STANDARD; PRT; 378 AA.  
AC P97368;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Homeobox protein Meis3 (Meis1-related protein 2).  
GN MEI3 OR MRG2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
CX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Swiss Webster;  
RX MEDLINE=97108670; PubMed=8950991;  
RA Nakamura T., Jenkins N.A., Copeland N.G.;  
RT Identification of a new family of Pbx-related homeobox genes.;  
RL Oncogene 13:2235-2242(1996).  
CC -! SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -! TISSUE SPECIFICITY: Expressed at high levels in the brain.  
CC Significant expression also observed in the heart, spleen and lung.  
CC -! DEVELOPMENTAL STAGE: Not expressed until 11 days in embryonic development.  
CC -! SIMILARITY: Belongs to the TALE/MEIS homeobox family.  
CC -! SIMILARITY: Contains 1 homeobox domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U57344; AAC52949.1; --  
DR HSP; P40424; I372  
DR TRANSPAC; T03406; --  
DR MGD; MG1108519; MrG2.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 1.  
DR ProDom; PD000010; Homeobox; 1.  
DR SMART; SMC0389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; FALSE\_NEG.  
DR PROSITE; PS50071; HOMEBOX\_2; 1.  
KW DNA-binding; Nuclear protein; Homeobox.  
FT DOMAIN 184 232 SER/THR-RICH.  
FT DOMAIN 234 262 ASP/GLU-RICH (ACIDIC).  
FT DNA BIND 265 327 HOMEBOX (TALE-TYPE).  
SQ SEQUENCE 378 AA; 41756 MW; 2175D0F4341E2484 CRC64;  
Alignment Scores:  
Pred. NO.: 1.77e-07 Length: 378  
Score: 194.00 Matches: 103  
Percent Similarity: 34.38% Conservative: 50  
Best Local Similarity: 23.15% Mismatches: 137  
Query Match: 5.91% Indels: 155  
DB: 15 Gaps: 15  
US-09-423-575-1 (1-1986) x MEI3\_MOUSE (1-378)  
QY 598 GATCATTTAAGACATCATGATGATCTCTCAACACACATGTGGGTTTGAAGCAATAGT 657  
Db 6 AspGluLeuArgHisTyrProGlyIleThrGluHisThrThrAlaLeuAlaSerPheSer 25  
QY 658 GAGTTTCAGGCATTTTCAGGTGTAGTTGGTCCCAAGTGACCAATGATG----- 705  
Db 26 Glu-----AlaAlaProSerValProArgAlaProGlyProTyr 38

QY 705 ----- 705  
Db 39 ThrProHisArgProProGlnLeuGlnAlaProGlyLeuAspSerLeuLysArg 58  
QY 706 -----TCTACATTGGTGAAGAGATTCCCGTCTTCTTAATTTTCGAATAAAGAAC 756  
Db 59 GluLysAspAspIleTyrGlyHisProLeuPheProLeuLeuAlaLeuValPheGluLys 78  
QY 757 AATGAGCTT-----TCATTGAGTCTTCATCA----- 783  
Db 79 CysGluLeuAlaThrCysSerProArgAspGlyAlaSerAlaGlyLeuGlySerProPro 98  
QY 784 -----GATGTT-----TCTGATGAATGCTCGAGATAAGTCTTTGTGACGATCAAGA 831  
Db 99 GlyGlyAspValCysSerSerAspSerPheAsnGluAspIleAlaPheAlaLysGln 118  
QY 832 TTAGCTTCAGACCAAGCTTCTTCAGCAGCAAGACATTCTTAATTAACGTTGTTACTCAA 891  
Db 119 IleArgSerGluArgProLeuPheSerSerAsnProGluLeuAspAsnLeuMetValGln 138  
QY 892 GGTTCCTCTCAACTTATTTGGCTCAAAATACCTTCACTCGTTCAGAAATACATCTCT 951  
Db 139 AlaIleGlnValLeuArgPheHisLeuLeuGluLeuGluLysValHisAspLeuCysAsp 158  
QY 952 CATTTCCGCCCATCTCGCTCGATTATTCATCTCGAGAACCGAGTCAGAGCTGCTAGT 1011  
Db 159 AsnPheCys----- 161  
QY 1012 TCAGCCTTACTTCACGTTTGGAGATATAACTGAGTTTCTTCATGCTGATCTTAATAAC 1071  
Db 161 ----- 161  
QY 1072 TCGAGCGCGGTTTCGGATCTACATTTCAAAGGAGAGCATTAGAAGCAAGAAACCCAT 1131  
Db 162 -----HisArgTyrIleThrCysLeuLysGlyLys 171  
QY 1132 CTC---TTGATCTTCTCAATGCTGGATGATCGATATAGTCATTCGCTAGATGAGATT 1188  
Db 172 MetProIleAspLeu-----ValIleGluAspArgAspGlySerCysArgGluAspLeu 189  
QY 1189 -----CATACGGTTATATACAGGTTTC 1209  
Db 190 GluAspTyrThrAlaSerCysProSerLeuProAspGlnAsnThrThr-TripIleArgAs 209  
QY 1210 CATGCTGCACCGAGTTAGATCCACAGTTACACCCGTTTTCCTCCCAACCGTTTCC 1269  
Db 209 pHisGluAspSerGlySerValHisLeuGlyThrProGly---ProSer-SerGlyGlyL 228  
QY 1270 TTCTTTATACAGAACCTGAGAGAGAATCTGCAAGAGATATCTCTATGGGATCTGTA 1329  
Db 228 euAlaSerGlnSerGlyAspAsnSerSerAspGlnGlyAsp----- 241  
QY 1330 TTGGAGAGAGGCAAGACAGACTCAAGAAACCTCTATGTTCCACCGATTGCTTCTT 1389  
Db 242 -----GlyLeuAspThrSerValAlaSerProSerS 252  
QY 1390 CAGCAGCTGAACGAAGAACCATCATGATTGAGACCTCAACGAGGT----- 1437  
Db 252 erAlaGlyGlu---AspGluAspLeuAspLeuGluArg-ArgArgAsnLysLysArgGly 270  
QY 1438 ---TTGCTCTGAGAAATCTGTTTCGGTCTACGAAATGGATGTTCCAAAACCTCCCTTCAC 1494  
Db 271 IlePheProLysValAlaThrAsnIleMetArgAlaThrLeuPheGlnHisLeuSerHis 290  
QY 1495 CTTTACCCGAAGAGATTCGGAGAAACATCTTCTAGCTATACGAAGTGGCTGACAGAAAT 1554  
Db 291 ProTyrProSerGluGluGlnLysLysGlnLeuAlaGlnAspThrGlyLeuThrIleLeu 310  
QY 1555 CAGGTATCAACCTGGTTTATAATGCGCGGTTAGGCTATGGAACCGGATGATAGAGAG 1614  
Db 311 GlnValAsnAsnTrpPheIleAsnAlaArgArgIleValGlnProMetIleAspGln 330  
QY 1615 ATGTATCGCGGAATGAACAGAGGAAGCTCAATAACAGTCACTTCAACCCACCGACCA 1674

Db 331 -----SerAsnArgThrGlyGlnGlyAlaSerPheAsnProGluGlyGln 345  
QY 1675 -----ACTCTCGAATGCCAAATCTGTT 1698  
Db 346 ProMetAlaGlyPheThrGluThrGlnProGlnValThrValArgThrProGlySerMet 365  
QY 1699 ATGATGACC 1707  
Db 366 GlyMetAsn 368  
RESULT 9  
MEI1\_XENLA  
ID MEI1\_XENLA STANDARD; PRT; 390 AA.  
AC P79937; P79938;  
DT 15-JUL-1998 (rel. 36, Created)  
DT 15-JUL-1998 (rel. 36, Last sequence update)  
DE 10-OCT-2003 (rel. 42, Last annotation update)  
DE Homeobox protein Meis1 (XMeis1).  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
OC xenopodinae; xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS MEIS1-1 AND MEIS1-2).  
RX MEDLINE=97202105; PubMed=9049632;  
RA Steelman S., Moskow J.J., Muzynski K., North C., Druck T.,  
RA Montgomery J.C., Huebner K., Daar I.O., Buchberg A.M.;  
RT "Identification of a conserved family of Meis1-related homeobox  
genes";  
RL Genome Res. 7:142-156(1997).  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event-Alternative splicing; Named isoforms=2;  
CC Comment=Additional isoforms seem to exist;  
CC Name=Meis1-1;  
CC IsoId=P79937-1; Sequence=Displayed;  
CC Name=Meis1-2;  
CC IsoId=P79937-2; Sequence=VSP\_002241;  
CC -!- SIMILARITY: Belongs to the TALE/MEIS homeobox family.  
CC -!- SIMILARITY: Contains 1 homeobox domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; U68386; AAB19196.1; -;  
DR EMBL; U68387; AAB19197.1; -;  
DR HSP; F41778; 1DUG.  
DR TRANSFAC; T03390; -;  
DR TRANSFAC; T03391; -;  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 1.  
DR ProDom; PDC00010; Homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; FALSE\_NEG.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
KW DNA-binding; Nuclear protein; Homeobox; Alternative splicing.  
FT DOMAIN 194 -240 SER/THR-RICH (ACIDIC).  
FT DOMAIN 242 -269 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 262 -267 POLY-ASP.  
FT DNA\_BIND 272 -334 HOMEBOX (TALE-TYPE).  
FT VARSPIC 373 -390 PMSGGMNMGMEGGWHYM -> LQMPPGSYISPSGPMGMSM  
FT AOPSVTSOMPLHQAOLRHGTSVHTYIPGHHHPAMMHGG  
FT PPOGMPISASSPSVINTGDFSMGSHVNNIHAQ (in  
FT isoform Meis1-2).  
FT /FTId=VSP\_002241.  
FT K -> Q (IN MEIS1-2).  
FT VARIANT 34 34



DR EMBL; AJ417081; CAd01142.1; -  
 DR EMBL; AB065001; BAB83665.1; -  
 DR TRANSFAC; T05155; -  
 DR Genew; HGNC:16714; PKNOX2.  
 DR InterPro; IPR001356; Homeobox.  
 DR Pfam; PF00046; homeobox; 1.  
 DR ProDom; PD000010; Homeobox; 1.  
 DR SMART; SM00389; HOX; 1.  
 DR PROSITE; PS00027; HOMEBOX\_1; FALSE\_NEG.  
 DR PROSITE; PS00071; HOMEBOX\_2; 1.  
 KW Homeobox; DNA-binding; Nuclear protein.  
 FT DNA\_BIND 279 338 HOMEBOX.  
 FT DOMAIN 8 11 POLY-PRO.  
 FT DOMAIN 425 439 POLY-GLU.  
 SQ SEQUENCE 460 AA; 50791 MW; AD2553BC17D9D7B0 CRC64;

Alignment Scores:  
 Pred. No.: 7,12e-07 Length: 460  
 Score: 186.50 Matches: 78  
 Percent Similarity: 41.99% Conservative: 74  
 Best Local Similarity: 21.55% Mismatches: 125  
 Query Match: 5.68% Indels: 85  
 DB: 1 Gaps: 16

US-09-423-575-1 (1-1886) x PNX2\_HUMAN (1-460)

QY 811 AGTCTTGTGTCAGCTACAAGATTAGCTCAGCAAGCTTCTTGACAGCAAGACATT 870  
 |||||  
 Db 26 SerLysAlaGlnAlaValHisSerAlaProSerAlaAlaAlaSerThrProValPro 45  
 |||||

QY 871 TCTAATAAGCTTGTACTCAAGTTTCTCTCAACTATATTGGCTCAAAATACCTTCAC 930  
 |||||  
 Db 46 SerAlaProIleAspProGlnAlaGlnLeuGluAlaAspLysArgAlaValTyrArgHis 65  
 |||||

QY 931 TCTGTTCAAGAAATATCTACTCTTCAGCTTTTCTTCACTTATTTGGCTCAAAATACCTTCAC 990  
 |||||  
 Db 66 ProLeuPheProLeuLeuThr-----JeuLeuPheGluLysCysGluGlnAlaThrGlnGly 84  
 |||||

QY 991 ACCGATCAGGAGCTGCTAGTTCAGCTTTTCTTCACTTATTTGGCTCAAAATACCTTCAC 1050  
 |||||  
 Db 85 SerGluCys--IleThrSerAlaSerPheAspValAspIleGluAsn-----Phe 100  
 |||||

QY 1051 CTTGATGGTGATCTCTAATACTCGGAGGGGGTTTCGGA-----TCTACA 1095  
 |||||  
 Db 101 ValHisGlnGlnGlnGlnGlnHisLysProPheSerAspAspProGluLeuAspAsn 120  
 |||||

QY 1096 TTTCAAAGAGAGACATTAGAAGCAAAAGAAACCATCTCTTGATCTTCTCAAAATGGTG 1155  
 |||||  
 Db 121 LeuMetValLysAlaIleGlnValLeuArgIleHisLeuLeuGluLeuGluLysValAsn 140  
 |||||

QY 1156 GAT-----GATCATATAGTCATTGGCTA----- 1179  
 |||||  
 Db 141 GluLeuCysLysAspPheCysAsnArgTyrIleThrCysLeuLysThrLysMethHisSer 160  
 |||||

QY 1180 ---GATGAGATTCATACGGTTTATATCAGCGTTCCATCGTCAACCGAGTAGATCCACAG 1236  
 |||||  
 Db 161 AspAsnLeuLeuArgAsnAspLeuGlyGlyProTyrSerProAsnGlnProSerLeuAsn 180  
 |||||

QY 1237 TTACACACCCGGTTTCCCTCCAA-----ACGTTTCTCTTATACAGAAC 1284  
 |||||  
 Db 181 LeuHisSerGlnAspLeuLeuGlnAsnSerProAsnSerMetSerGlyValSerAsnAsn 200  
 |||||

QY 1285 -----CTGAGAGAGAGA 1296  
 |||||

Db 201 ProGlnGlyIleValValProAlaSerAlaLeuGlnGlnGlyAsnIleAlaMetThrThr 220  
 |||||

QY 1297 ATCTGCAAGAGATAATCTCTATGGGATCT-----GTATGGGAG 1335  
 |||||  
 Db 221 ValAsnSerGlnValValSerGlyGlyAlaLeuTyrGlnProValThrMetValThrSer 240  
 |||||

QY 1336 AGAGGCAACACAGACTCAAGAA----- 1359  
 |||||

Db 241 GlnGlyGlnValValThrGlnAlaIleProGlnGlnGlyAlaIleGlnIleGlnAsnThrGln 260  
 |||||

QY 1360 -----ACCTCTATGTTCCACGACATTGCCTTCTTCTCAGCAGCTGAAACGA 1404  
 |||||  
 Db 261 ValAsnLeuAspLeuThrSerLeuLeuAspAsn-----GluAspLysLysSer 276  
 |||||

QY 1405 AAGAACCATCAGATTTCGAGACCTCAACGAGGT---TTGCCTGAGAAATCTGTTTCGTT 1461  
 |||||  
 Db 277 LysAsn-----LysArgGlyValLeuProLysHisAlaThrAsnIle 290  
 |||||

QY 1462 CTACGGAATTCGATGTTCCAAACTTCTTCCACCTTACCCGAAAGATTCCGAGAAACAT 1521  
 |||||  
 Db 291 MetArgSerTrpLeuPheGlnHisLeuMethHisProTyrProThrGluAspGluLysArg 310  
 |||||

QY 1522 CTTCTAGCTATACGAAGTGGCTTCACAAGAGTTCAGGTTCAGACTCGTTTATTAATGCG 1581  
 |||||  
 Db 311 GlnIleAlaAlaGlnThrAsnLeuLeuGlnValAsnAsnTrpPheIleAsnAla 330  
 |||||

QY 1582 CGGTTTAGCTTAGGCTAGAGCCGATGATAGAGAGATGATGCGGAA-----ATGAACAAG 1635  
 |||||  
 Db 331 ArgArgArgIleLeuGlnProMetLeuAspAlaSerAsnProAspProAlaProLysAla 350  
 |||||

QY 1636 AGAAGCTCAATATACAGTCACATTCAACCCACGACCACTCTTCAATG---CCAAAA 1692  
 |||||  
 Db 351 LysLysIleLysSerGlnHis-----ArgProThrGlnArgPheTrpProAsn 366  
 |||||

QY 1693 TCTGTT 1698  
 |||||  
 Db 367 SerIle 368  
 |||||

RESULT 11  
 MEI3\_HUMAN  
 ID MEI3\_HUMAN STANDARD; PRT; 382 AA.  
 AC Q99687; Q9NPW2;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Homeobox protein Meis3 (Meis1-related protein 2) (Fragment).  
 GN MEIS3 OR MRG2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Blum H., Bauersachs S., Mewes H.-W., Weil B., Wiemann S.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 FN [2]  
 RP SEQUENCE OF 175-382 FROM N.A.  
 RX MEDLINE=97202105; PubMed=9049632;  
 RA Steelman S., Moskow J.J., Muzynski K., North C., Druck T.,  
 RA Montgomery J.C., Huebner K., Daar I.O., Buchberg A.M.;  
 RT "Identification of a conserved family of Meis1-related homeobox  
 genes";  
 RL Genome Res. 7:142-156 (1997).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- SIMILARITY: Belongs to the TALE/MEIS homeobox family.  
 CC -!- SIMILARITY: Contains 1 homeobox domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/ebis-sib.ch/announcement/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; AL359938; CAB95771.1; -  
 CC EMBL; U68385; AAB19195.1; -  
 CC HSRP; P40424; 1B72.  
 CC TRANSFAC; T03412; -  
 CC Genew; HGNC:7002; MEIS3.  
 CC GO; GO:0005634; C:nucleus; IBS.











```

Db      291 GlnlyseuAlaIeuAlaGluSerThrGlyLeuAspGlnlysglnIleAsnAsnTrpPhe 310
QY      1573 ATAATATCGCGGGTGTAGCTGATGAAGCCGATGATGAACAGATGTATGCAGAAATCAAC 1632
Db      311 IleAsnGlnArglylsArgHISrTpLyProSerGluaspMetGlnPheValMetAsp 330

RESULT 14
HKU3_ARATH STANDARD; PRT; 431 AA.
ID     HKU3_ARATH          ID     AC
AC      P48000;
DT      01-FEB-1996 (Rel. 33, Created)
DT      01-FEB-1996 (Rel. 33, Last sequence update)
DE      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Homeobox protein knotted1 like 3 (KNAT3).
GN      KNAT3 OR AT5G25220 OR F2J36.i18.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC      euroids II; Brassicales; Brassicaceae; Arabidopsais.
ON      NCBI_TaxID=3702;
RX      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97304677; PubMed=9161040;
RX      Serikawa K-A., Martinez-Laborda A., Kim H.S., Zambryski P.C.;
RX      "Localization of expression of KNAT3, a class 2 knotted1-like gene.";
RX      Plant J. 11:853-861(1997).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      STRAIN=cv. Columbia;
RX      MEDLINE=21016721; PubMed=11130714;
RX      Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RX      Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashina K.,
RX      Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RX      Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,
RX      Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide R.,
RX      Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RX      Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RX      Stoneking T., Pupin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RX      Belter B., Cordum H., Cordes M., Courtney L., Courtney W., Dante P.,
RX      Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RX      Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,
RX      Wagner-McPherson C., Wollam A., Yokum M., Bell M., Dedhia N.,
RX      Farnell L., Shah R., Rodriguez M., Hoan See L., Vil D., Baker J.,
RX      Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,
RX      Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RX      Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RX      Entian K.-D., Terry N., Hartley N., Bent E., Johnson S.,
RX      Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
RX      Rampeger U., Wedler H., Balke K., Wedler E., Peters S.,
RX      van Staveren M., Dirkse W., Mooijman P., Klein Lankhorst R.,
RX      Weizensegger T., Bothe G., Rose M., Hauf J., Berneriser S., Hempel S.,
RX      Feldpausch M., Lamberth S., Villarreal R., Gielen J., Ardiles W.,
RX      Benus O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H.,
RX      Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.,
RX      "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RX      thaliana.",
RX      Nature 408:823-826(2000).
RL      Nature 408:823-826(2000).
CC      -!- SUBCELLULAR LOCATION: Nuclear (probable).
CC      -!- SIMILARITY: Belongs to the TAIR/KNOX homeobox family.
CC      -!- SIMILARITY: Contains 1 homeobox domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; X92392; AAC63130.1; -.
CC      DR      EMBL; AC006259; AAC98441.1; -.
CC      DR      TRANSFAC; TC04044; -.
CC      DR      InterPro; IPRO05539; ELK.

```



Search completed: September 2, 2004, 12:07:35  
Job time : 41 secs

```
Db      26 SerSerProLeuSerAlaValIleSerProProGlnGlnGlnGlnHisGlnGln 45
QY      721 GAAGATTTCCTCGTTCTTAATTCGAATAAAGAAACAATGAGCTTTCATTGAGTCTTGCA 780
Db      46 GinSerAlaGlyTyrLeu-----AlaHisSerProLeuSerLeuAsnThrAla 61
QY      781 ---TCAGATGTTTCT-----GATGAATGCTCGAGATAAGTCTTTGTGCAGCT 825
Db      62 ProProGlyValSerHisGlyGlySerGlyCysSerAsnProValLeuGlnLeuAla 81
QY      826 ACAAGATTAGCCTCAGACCAAGCTTCTTCAGCAGCAAGACATTTCTAATAACGTTGTT 885
Db      82 AsnGlySerLeuLeuGlnAlaCysAlaLysAlaAlaLysGluProSerSer----- 98
QY      886 ACTCAAGGTTTCTCTCAACTTATATTGGCTCAAAATACCTTCACTCTCTTCAA----- 939
Db      99 -----SerSerTyrAlaAlaAspValGluAlaIle 108
QY      940 -----GAAATACTATCTCAT-----TTGCGCGCATACTCGCTCGAT 975
Db      109 LysAlaLysIleLeSerHisProHisTyrSerSerLeuLeuAlaIleTyr---LeuAsp 127
QY      976 TATTCACTCTCGAGAACCGAGTCAGGAGCTGTAGTTCAGCCTTTACTTCACTGTTTGAG 1035
Db      128 CysGlnLysValGly-----AlaProProGluValSerAlaArgLeuThr 142
QY      1036 AATATAACTGAGTTCTTGATGTGATCTTAATACTCGAGCGCGGTTTCGATCTACA 1095
Db      143 AlaValAlaGlnAspLeuGluLeuArgGlnArgThrAlaLeuGlyGlyLeuGlyThrAla 162
QY      1096 TTTCAAAGGAGACATTAGAACAAAGAAACCCATCTCTTGATCTCTTCAAATGGTG 1155
Db      163 ThrGlu----- 164
QY      1156 GATGATCGATATAGTCATTGCGTAGATGAGATTACGCTTATATACGCTTCCATGCT 1215
Db      165 -----ProGluLeuAspGlnPheMetGluAlaTyrHisGlu 176
QY      1216 GCAACCGAGTTAGTCCAGTTACACACCGGTTTGCCCTCCAAACCGTTTCTCTTA 1275
Db      177 MetLeuValLysTyrArgGluLeuThrArgProLeuGlnGluAlaMetGluPheLeu 196
QY      1276 TACAAG-----AACCTGAGAGAGAGATCTGCAAGAAGATA 1311
Db      197 ArgArgValGluThrGlnLeuAsnSerLeuSerIleSerGlyArgSerLeuArgAsnIle 216
QY      1312 ATCTCTATCGGATCTGTATTGAGAGAGCGCAAGACAGACTCAAGAAACCTCTATG--- 1368
Db      217 LeuSerThrGlySerSerGluGluAspGlnGluGlySerGlyGlyGluThrGluLeuPro 236
QY      1369 ---TTCCACCAGATTGCCCTTCTTCAGCAGCTGAAACGAAACCAACATCAGATTGGAGA 1425
Db      237 GluIleAspAlaHisGlyValAspGlnGluLeuLys-----HisHisLeuLeuLys 253
QY      1426 CCTCAACGAGGT----- 1437
Db      254 LysTyrSerGlyTyrLeuSerSerLeuLysGlnGluLeuSerLysLysLysGly 273
QY      1438 ---TTGCTTGAGAAATCTGTTTCGGTTCTACGGAATTGATGTTCCAAACCTTCTTCAAC 1494
Db      274 LysLeuProLysGlnAlaArgGlnGlnLeuLeuSerTrpTrpGluMetHisTyrLysTrp 293
QY      1495 CTTTACCCGAAAGATTCCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAGAGT 1554
Db      294 ProTyrProSerGlnSerGlnLysValAlaLeuAlaGluSerThrGlyLeuAspLeuLys 313
QY      1555 CAGGTATCAAACTGGTTTATAATGCGCGGTTAGGCTATGGAACCGCATGATGAGAAG 1614
Db      314 GlnIleAsnAsnTrpPheIleAsnGlnArgLysArgHisTrpLysProThrAspGluMet 333
QY      1615 ATGTATCGCGAAATGAAC 1632
Db      334 GlnPheValMetMetAsp 339
```

mis Page Blank (uspto)

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 2, 2004, 11:35:18 ; Search time 118 Seconds  
(without alignments)

10085.887 Million cell updates/sec

Title: US-09-423-575-1

Perfect score: 384  
Sequence: 1 attagttataaatgttc.....ttaaaaaaaaaaaaaaaaaa 1886

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\*\_n2p.model -DRV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09423575/runat\_01092004\_161106\_18570/app\_query.fasta\_1.2055  
-DB=SPREMBL\_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09423575.@CGN\_1\_162.@runat\_01092004\_161106\_18570 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCC=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPREMBL\_25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriopl:\*

17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2453	74.7	469	10 Q7X9S7	Q7X9S7 arabidopsis

ID	Q7X9S7	PRELIMINARY;	PRT;	469 AA.
AC	Q7X9S7			
DT	01-OCT-2003 (TRENBLrel. 25, Created)			
DT	01-OCT-2003 (TRENBLrel. 25, Last sequence update)			
DT	01-OCT-2003 (TRENBLrel. 25, Last annotation update)			
DE	H1 gene protein.			
GN	H1 GENE.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV, Colombia; TISSUE=Rosette leaves;			
RX	MEDLINE=95210927; PubMed=7696878;			
RA	Quaedvlieg N., Dockx J., Rook F., Weisbeek P., Smeekens S.;			
RT	"The Homeobox gene AtH1 of Arabidopsis is depressed in the			
RT	photomorphogenic mutants cop1 and get1."			
RL	Plant Cell 7:117-129(1995).			
DR	EMBL; X80127; CAC51426.1; -			
SQ	SEQUENCE 469 AA; 53477 MW; 3DC7C6A976819574 C			

## ALIGNMENTS

RESULT 1

Q7X9S7

ID Q7X9S7 PRELIMINARY; PRT; 469 AA.

AC Q7X9S7

DT 01-OCT-2003 (TRENBLrel. 25, Created)

DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE H1 gene protein.

GN H1 GENE.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosid II; Brassicales; Brassicaceae; Arabidopsi

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV, Colombia; TISSUE=Rosette leaves;

RX MEDLINE=95210927; PubMed=7696878;

RA Quaedvlieg N., Dockx J., Rook F., Weisbeek P., Smeekens S.;

RT "The Homeobox gene AtH1 of Arabidopsis is depressed in the

photomorphogenic mutants cop1 and get1."

RL Plant Cell 7:117-129(1995).

DR EMBL; X80127; CAC51426.1; -

SQ SEQUENCE 469 AA; 53477 MW; 3DC7C6A976819574 C

## Alignment Scores:

Pred. No.: 9-9e-208 Length: 469  
 Score: 2453.00 Matches: 469  
 Percent Similarity: 99.15% Conservative: 0  
 Best Local Similarity: 99.15% Mismatches: 0  
 Query Match: 74.70% Indels: 4  
 DB: 10 Gaps: 1

US-09-423-575-1 (1-1886) x Q7X9S7 (1-469)

QY 304 ATGCAACAACAACAACAACAACACTTTTCTCTGGATAAATGTCATGACTAACAA 363  
 Db 1 MetAspAsnAsnAsnAsnAsnThrPheSerSerLeuAspAsnValMetThrAsnGln 20  
 QY 364 AATCCTCTTCTCATGGATTATACCTTCAAGAGAGATTCAACTTCATTCTCAACAATG 423  
 Db 21 AsnProLeuLeuMetAspPheIleProSerArgGluAsp-----SerThrMet 36  
 QY 424 CTTCATGGAATACCATCATGATCAGATCCTCTACAAATGGGTGGCTTTGCATATTTCAAT 483  
 Db 37 LeuProTrpAsnThrIleArgSerAspProLeuGlnMetGlyGlyPheAspIlePheAsn 56  
 QY 484 TCTATCTGACTAAACAATACTTATCATCTCTCCACGGTCTATCGATGTTCAAGATAAC 543  
 Db 57 SerMetLeuThrAsnLysTyrLeuSerSerSerProArgSerIleAspValGlnAspAsn 76  
 QY 544 CGCAATGTTGAGTTCAATGCT 603  
 Db 77 ArgAsnValGluPheMetAlaProProHisProProLeuHisProLeuAspHis 96  
 QY 604 TTAAGACACTGATGATTCCTCAACAACATGTTGGGTTTTGAAGCAAAATAGTGAGTTT 663  
 Db 97 LeuArgHisTyrAspAspSerSerAsnAsnMetTrpGlyPheGluAlaAsnSerGluPhe 116  
 QY 664 CAGGCATTTTCAGGTGATCTGGTCCAACTGAACCAATGATCTACATTCGGTGAAGAA 723  
 Db 117 GlnAlaPheSerGlyValValGlyProSerGluProMetMetSerThrPheGlyGluGlu 136  
 QY 724 GATTTCCCGTTCTAATTTTGAATAAAGAAACATGAGCTTTTCATTCAGTCTTCATCA 783  
 Db 137 AspPheProPheLeuIleSerAsnLysArgAsnAsnGluLeuSerLeuSerLeuAlaSer 156  
 QY 784 GATGTTTCTGATGATGTCGGAGATAAGTCTTTTGTGCAGCTACAGATTAGCCTTCAG 843  
 Db 157 AspValSerAspGluCysSerGluIleSerLeuCysAlaAlaThrArgLeuAlaSerGlu 176  
 QY 844 CAGCTTCTGAGCAGCAGCAACATCTTCTATACAGTTGTGTACTCAAGGTTCTCTCAA 903  
 Db 177 GlnAlaSerCysSerSerLysAspIleSerAsnAsnValValThrGlnGlyPheSerGln 196  
 QY 904 CTATATTTGGCTCAAAATACCTTCACTCTGTTCAGAAATACTATCTCATTTCCGCCCA 963  
 Db 197 LeuIlePheGlySerLysTyrLeuHisSerValGlnGluLeuSerHisPheAlaAla 216  
 QY 964 TACTGCTCGATATTCATCTCAGAGAACCGAGTACAGAGCTGCTAGTTTCAGCCTTACT 1023  
 Db 217 TyrSerLeuAspTyrSerSerArgGlyThrGluSerGlyAlaAlaSerSerAlaPheThr 236  
 QY 1024 TCACGTTTTGAGATATTAATCTCAGTTCTTGTGTGATTTCTTAATCTCGAGGCGGGT 1083  
 Db 237 SerArgPheGluAsnIleThrGluPheLeuAspGlyAspSerAsnAsnSerGluAlaGly 256  
 QY 1084 TTCGATCTACATTTCAAGAGAGACATTAGAGCAAAAGAAACCCATCTCTTGGATCTT 1143  
 Db 257 PheGlySerThrPheGlnArgAlaLeuGluAlaLysLysThrHisLeuLeuAspLeu 276  
 QY 1144 CTTCAAATCGTGATCATGATATAGTATTCGCTAGATGATTCATACGTTATATCA 1203  
 Db 277 LeuGlnMetValAspAspArgTyrSerHisCysValAspGluIleHisThrValIleSer 296  
 QY 1204 GGTTCCTCATGTCGAACCGAGTTAGATCCACAGTTTACACCCCGGTTTCCCTCCCAACC 1263  
 Db 297 AlaPheHisAlaAlaThrGluLeuAspProGlnLeuHisThrArgPheAlaLeuGlnThr 316

## RESULT 2

QY026 PRELIMINARY; PRT; 323 AA.  
 AC QY026;  
 DT 01-OCT-2003 (T-EMBLrel. 25, Created)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE Bell-like homeodomain protein 4 (fragment).  
 GN Bt4.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 OC Lamiales; Solanales; Solanaceae; Solanum.  
 OX NCBI\_TaxID=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ron N., Parnis A., Lifschitz E.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF375967; AAP47026.1; -;  
 KW DNA-binding; Homeobox; Nuclear protein.  
 FT NON TER 1 1  
 SQ SEQUENCE 323 AA; 37216 MW; DED2C654B526D6F1 CRC64;

Alignment Scores:  
 Pred. No.: 3-37e-56 Length: 323  
 Score: 736.50 Matches: 166  
 Percent Similarity: 65.16% Conservative: 36  
 Best Local Similarity: 53.55% Mismatches: 72  
 Query Match: 22.43% Indels: 37  
 DB: 10 Gaps: 7

US-09-423-575-1 (1-1886) x QY026 (1-323)

QY 793 GATGAATGTCGAGATAAGTCTTCTGTCAGCTACA-----AGA 831  
 Db 2 AspGlnCysAspIleSerCysSerGlyValThrAsnHisAlaPheProGlnArg 21  
 QY 832 TTAGCTTCAGAGCAAGCTTCTTGCAGCAGCAAGACATTCT-----AATAACGTT 882  
 Db 22 PheAspSerGluLeuThrSerCysAsnSerArgAsnLeuSerLeuSerPheGlySerTyr 41







Db 192 eArgGlyThrLeuAspGlySerSerSerAsnMetValLeuGlySerLysTyrLeuLysAl 212  
 QY 933 TTTTCAAGAAATATCTATCTCATTTCCGCCGATCTGCTGCTGATTATTCATCTCGAGAAC 992  
 Db 212 aAlaGlnGluLeuLeuAspGlu---ValValAsnIleValGlyLysSerIleLysGlyAs 231  
 QY 993 CGAGTCAGGAGCTGCTAGTCTTACGCTTTTACCTTACGCTTTTGAATATAACTGATTTCT 1052  
 Db 231 pAspGlnLysLysAspAsnSer-----MetAsnLysGluSerMetProLeuAlaSe 248  
 QY 1053 TGATGCTGATTTCAATACTCGAGCGGGGTTTCGATCT----- 1092  
 Db 248 aAspValAsnThrAsnSerSerGlyGlyGlyGluSerSerAspGlnLysAsnGluVa 268  
 QY 1093 -----ACATTTCAAGAGAGAGCATTAGAACAAGAAACCCATCTCTT 1136  
 Db 268 lAlaValGluLeuThrThrAlaGlnArgGlnGluLeuGlnMetLysLysAlaLysLeuLe 288  
 QY 1137 GGATCTCTTCAAAATGGTGGATGATCATATAGTCATTCGCTAGATGATCATACGCT 1196  
 Db 288 uAlaMetLeuGluGluValGluGlnArgTyrArgGlnTyrHisHisGlnMetGlnIleIl 308  
 QY 1197 TATATCAGCGTTCATCTGTCGACCGAGTTA---GATCCACAGTTACACACCGGTTTGC 1253  
 Db 308 eValLeuSerPheGluGlnValAlaGlyIleGlySerAlaLysSerTyrThrGlnLeuAl 328  
 QY 1254 CTCCAAACCGTTTCTCTTATACAAACCTGAGAGAGAACTGCAAGAGATAAT 1313  
 Db 328 aleuHisAlaIleSerLysGlnPheArgCysLeuLysAspAlaIleAlaGluGlnVally 348  
 QY 1314 CTCT-----ATGGGATCTGTATTGGAGAGGCA 1343  
 Db 348 sAlaThrSerLysSerLeuGlyGluGluGlyLeuGlyGlyLysIleGluGlySerAr 368  
 QY 1344 AGACAGAGCTCAAGAACTCTATGTTCCACGAGATTCCTTCTTCAGCAGCTGAACG 1403  
 Db 368 gLeuLysPheValAspHisHisLeuArgGlnGlnArgAla---LeuGlnGlnIleGlyMe 387  
 QY 1404 AAGAACCACATCAGATTGGAGACCTCAACGAGTTTGCCTGAGAAATCTGTTTCGGTCT 1463  
 Db 387 tMetGlnProAsnAlaTrpArgProGlnArgGlyLeuProGluArgAlaValSerValle 407  
 QY 1464 AGGAAATTGGATGTTCCAAACTTCTTCCACCTTACCCGAAAGATTTCGAGAAACATCT 1523  
 Db 407 uArgAlaTrpLeuPheGluHisPheLeuHisProTyrProLysAspSerAspLysIleMe 427  
 QY 1524 TCTAGCTATACGAAGTGGCTTGCACAAAGATCAGGTATCAAACTGTTTATAAATGCGCG 1583  
 Db 427 tLeuAlaLysGlnThrGlyLeuThrArgSerGlnValSerAsnTrpPheIleAsnAlaAr 447  
 QY 1584 GGTAGCTATGGAAGCCGATGATAGAGAGATGTATCGGAAATGAACAAGAGGAAGCT 1643  
 Db 447 gValArgLeuTrpLysProMetValGluGluMetTyrLeuGluGluValLysAsnGlnGl 467  
 QY 1644 CAATAACACT 1653  
 Db 467 uGlnAsnSer 470

RESULT 5

Q94KL4  
 ID Q94KL4 PRELIMINARY; PRT; 680 AA.  
 AC Q94KL4;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE BEL1-like homeodomain 1.  
 GN BLH1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotys; rosids;  
 OC eurooids 1; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID:3702;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA Pickowich M.S., Samach A., Modrusan Z., Haughn G.W.;  
 RT "A family of BEL1-like homeodomain (BLH) proteins in Arabidopsis  
 thaliana";  
 RL Submitted (PEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF353034; AAK43836.1; -  
 DR GO: GO:0005634; C:nucleus; IEA.  
 DR GO: GO:0003700; F:transcription factor activity; IEA.  
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro: IPR001356; Homeobox.  
 DR InterPro: IPR006563; POX.  
 DR Pfam: PF00046; homeobox; 1.  
 DR ProDom: PD00010; Homeobox; 1.  
 DR SMART: SM00359; HOX; 1.  
 DR SMART: SM00574; POX; 1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS00071; HOMEBOX\_2; 1.  
 KW DNA-binding; Homeobox; Nuclear protein.  
 SQ SEQUENCE 680 AA; 74457 MW; AA9E845082F5799C CRC64;

Alignment Scores:

Pred. No.: 9.28e-33 Length: 680  
 Score: 472.00 Matches: 152  
 Percent Similarity: 42.01% Conservative: 74  
 Best Local Similarity: 28.25% Mismatches: 193  
 Query Match: 14.37% Indels: 119  
 Gaps: 13

US-09-423-575-1 (1-1886) x Q94KL4 (1-680)

QY 283 ATAAAGTTTCATAGAAACCCATGGACACACACACACACACACTTTTGTCTCTG 342  
 Db 32 ValGlnTyrThrGlnGlnAspAspSerAsnAsnAsnAsnSerAsnSerAsn 51  
 QY 343 GAT-----AATGTCATGACTAACCAAAATCTCTTCATGGATTTTATACCTTCA 393  
 Db 52 AsnAsnAsnThrAsnThrAsnThrAsnAsnAsnSerSerPheValPheLeuAspSer 71  
 QY 394 AGAGAAGATTCAACTTCATCTCAACAATGCTTCCATGGAAATACCATCAGATCAGTCT 453  
 Db 72 HisAlaPro-GlnProAsnAlaSerGlnGlnPheValGlyLeuProLeu----- 87  
 QY 454 CTACAAATGGGTGGCTTTGATATTTTCAATCTATGCTGACTAACAAATACCTATCATCT 513  
 Db 88 -----SerGlyHis-- 90  
 QY 514 TCTCCCGGTCTATCGATGTTCAAGATAACCGCAATGTTGAGTTCATGGCTCTCTCTCT 573  
 Db 91 GluAlaAlaSerIleThrAlaAlaAspAsnIleSerValLeuHisGlyTyrProProArg 110  
 QY 574 CATCTCTCCATCTCATCTCTTGGATCATTTAAGACACTATGATGATCTCTCAACAAC 633  
 Db 111 ValGlnTyrSerLeuTyrGly-----SerHisGlnValAspProThrHisGln 126  
 QY 534 ATGTGGGGTTTGAAGCAAAATAGTAGTTCAGGCATTTTCAGGTAGTTGGTCCAAGT 693  
 Db 127 GlnAlaAlaCysGluThrProArgAlaGlnGlnGlyLeuSerLeuThrLeuSerSerGln 146  
 QY 694 GAACCAATGATGCTACATTCGGTGAAGAAGATTTCCTGTTTCTTAATTTGCAATAAAGA 753  
 Db 147 GlnGlnGlnGlnGlnGlnHisHisGlnGlnHisGlnProIleHisValGlyPheGlySer 166  
 QY 754 AACAAATGAGCTTCATTGAGTCTTGATTCAGATCTGATGATGATCTCGGAGATAAGT 813  
 Db 167 GlyHisGlyGluAspIleArgValGlySer----- 176  
 QY 814 CTTTGTGCGCTACAAGATTAGCTCAGAGCAAGCTTCTTTCAGCAGCAAGACAAAGATTTCT 873  
 Db 177 -----GlySerThr 179  
 QY 874 AATAACCTGTGTACTCAAGTTTCTCTCAACTTATATTTGGCTCAAAATACCTTCACTCT 933  
 Db 180 GlySerGlyValThrAsnGlyIleAlaAsnLeuVal---SerSerLysTyrLeuLysAla 198

```

QY 934 GTTCAGAAATACATCTCTTCCTCCGCGCATCTCGCTCGATTATTCATCTCGAGAAC 993
DB 199 AlaGlnGluLeuLeuAspGluValValAsnAlaAspSerAsp----- 212
QY 994 GAGTCAGGAGCTGAGTACCTTCACCTTTTACTTCACCTTTTTCAGAAATATA---ACTCAGATT 1050
DB 213 AspMetAsnAlaSerGlnLeuPheSerSerLysLysGlySerCysGlyAsnAspLys 232
QY 1051 CTTGATGGTATCTTAATTAACCTCGGAGCGGGTTTCGGATCT----- 1092
DB 233 ProValGlyGluSerSerAlaGlyAlaGlyGlyGluGlySerGlyGlyValaGluAla 252
QY 1093 -----ACATTCAAGAGAGAGATTAAGACCAAGAA 1125
DB 253 AlaGlyLysArgProValGluLeuGlyThrAlaGluArgGlnGluLeuMetLysLys 272
QY 1126 ACCATCTCTTGGATCTCTTCAATGGTGGATGATCATGATGATGATGATGATGATG 1185
DB 273 AlaLysLeuSerAsnMetLysHisGluValGluGlnArgTyrArgGlnTyrHisGlnGln 292
QY 1186 ATTCATACAGTTATATCAGCTTCCATGCTGCAACCGAGTTA---GATCCACAGTTACAC 1242
DB 293 MetGlnMetValLysSerPheGluGlnAlaAlaGlyLeuGlySerAlaLysSerTyr 312
QY 1243 ACCCGTTTGGCTCCAAACCGTTTCTTCTTATACAAGAACCTGAGAGAGAGATCTGC 1302
DB 313 ThrSerLeuAlaLeuLysThrLysSerArgGlnPheArgCysLeuLysGluAlaAla 332
QY 1303 AAGAGATATCTCTATGGATCTGTATGGAGAGAGCAAGACAACTCAGAAAC 1362
DB 333 -----GlyGlnLeuLysAlaAlaAsnLysSerLeuGlyGluGluAsp 346
QY 1363 TCTATG-----TTCCACGACGATGCTCT 1386
DB 347 SerValSerGlyValGlyArgPheGluGlySerArgLeuLysPheValAspHisLeu 366
QY 1387 CTTGACGAG-----CTGAACGAAGAACCATCAGATTGG 1422
DB 367 ArgGlnGlnArgAlaLeuGlnGlnLeuGlyMetIleGlnHisProSerAsnAlaTrp 386
QY 1423 AGACCTCAACGAGTTTGCTCAGAAATCTGTTTCGGTCTACGGAATGGATGTCACAA 1482
DB 387 ArgProGlnArgGlyLeuProGlnArgAlaValSerValLeuArgAlaTrpPheGlu 406
QY 1483 AACTTCTCTTCACTTACCGGAATTCGGAGAAATCTTCTAGCTATACGAAGTGGC 1542
DB 407 HisPheLeuHisProGlyProLysAspSerAspLysHisMetLeuAlaLysGlnThrGly 426
QY 1543 TTGACAGAGTCAGGTATCAACTGTTTATATGCGGGTATGAGCTATGGAAGCG 1602
DB 427 LeuThrArgSerGlnValSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysPro 446
QY 1603 ATGATAGAGAGATGTATCGGAAATGAACAGAGAAAGCTCAATAACAGTCACATTCAA 1662
DB 447 MetValGluGluMetTyrMetGluGluMetLysGluGlnAlaLysAsnMetGlySerMet 466
QY 1663 CCCAACGGACCA-----ACTCTTCGATG 1686
DB 467 GluLysThrProLeuAspGlnSerAsnGluAspSerAlaSerLysSerThrSerAsnGln 486
QY 1687 CCAAAATCTGTTATGATGACCAAGCAATGATCAATAAATAAGCAACAAT 1734
DB 487 GluLysSerProMetAlaAspThrAsnTyrHisMetAsnProAsnHis 502

RESULT 6
Q9SJ56 PRELIMINARY; PRT; 680 AA.
AC Q9SJ56;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative homeodomain transcription factor (At2g35940).

```

```

GN AT2G35940.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
PC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Rongming C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.E., Unayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768 (1999).
[2]
RN SEQUENCE FROM N.A.
PC STRAIN=cv. Columbia;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
RA Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene At2g35940 (GI:15227535).";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
RA Davis R.W., Ecker J.R., Theologis A.;
RT "Arabidopsis Open Reading Frame (ORF) Clones.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
[6]
RN SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Dale J.M., Hayashizaki Y., Hsuan V.W., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007017; AAD21463.1; -
DR EMBL; AY072175; AAL59997.1; -
DR EMBL; AY096548; AAM20198.1; -
DR EMBL; AY099854; AAM20705.1; -
DR EMBL; BT008422; AAP37781.1; -

```

DR	PIR; H84774; H84774.
DR	HSSP; P01366; IAKH.
DR	GO; GO:0005634; C:nucleus; IEA.
DR	GO; GO:0003700; F:transcription factor activity; IEA.
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR	InterPro; IP001356; Homeobox.
DR	Pfam; PF00045; homeobox_1.
DR	ProDom; PD00010; Homeobox; 1.
DR	PROSITE; PS00027; HOMEBOX_1; 1.
DR	PROSITE; PS00071; HOMEBOX_2; 1.
KW	DNA-binding; Homeobox; Nuclear protein.
SQ	SEQUENCE 680 AA; 74463 MW; 45COE77996AB8737 CRC64;

  

Alignment Scores:	
Pred. No.:	9.28e-33
Score:	472.00
Percent Similarity:	42.01%
Best Local Similarity:	28.25%
Query Match:	14.37%
DB:	1.0

  

US-09-423-575-1 (1-1886) x Q9SJ56 (1-680)	
QY	283 ATAAAGTTTCATAGAACCACCAATCGACAACAACAACAACAACACTTTTAGTCTCTG 342
Db	32 ValGlnfThrGlnGlnAsnAspSerAsnAsnAsnAsnAsnSerAsnAsnSerAsn 51
QY	343 GAT-----AATGTCAGACTAACCAAATCCTCTTCATGGATTATATACCTTCA 393
Db	52 AsnAsnAsnThrAsnThrAsnThrAsnAsnAsnAsnSerSerPheValPheLeuAspSer 71
QY	394 AGAAGAATTCAACTTCATTCACCAATGTCTCCATGGATAACCATCAGATCAGATCCT 453
Db	72 HisAlapro-GlnProAsnAlaserGlnGlnPheValGlyleProLeu----- 87
QY	454 CTCAAAATGGTGCGTTTGATATTTTCAATCTATGCTAGCTACAAACTACTTATCATCT 513
Db	88 -----SerGlyHis-- 90
QY	514 TCTCCACGGCTATPCGATGTTCAGAGTAACCGCAATGTGAGTTGCATGGCTCCTCCTCT 573
Db	91 GlualaalaserIleThrAlaalaspAsnIleserValleuHisGlyTyrrProProarg 110
QY	574 CATCTCCTCCATTCATCCCTTTGGATCATNTTAAGACACTANGATGATTOCTCAACAAC 633
Db	111 ValGlnfyrSerLeutyryGly-----SerHisGlnValaspProThrHisGln 126
QY	634 ATGTGGGGTTTTGAAGCAAAATAGTAGTTCAGGCATTTTCAGGTGATGTGGTCCAAGT 693
Db	127 GlnAlaalacysGluThrProargAlaGlnGlnGlyLeuSerLeuThrLeuSerSerGln 146
QY	694 GAACCAATGATGCTACATTCGGGTGAAGAAGATTTCCTGTTTCAATTTTGAATATAAGA 753
Db	147 GlnGlnGlnGlnGlnHisHisGlnGlnHisGlnProIleHisValGlyPheGlySer 166
QY	754 AACAAATGAGCTTTCATTGAGTCITGCATCAGATGTTTCTGATGAATGCTCGAGATAAGT 813
Db	167 GlyHisGlyGluaspIleArgValglyser----- 176
QY	814 CTTTGTGCAGCTACAAGATTAGCCTCAGACGAGCTTCTTGACAGCAAGACATTCT 873
Db	177 -----GlySerThr 179
QY	874 AATAAGCTTGTTACTCAAGTTTCTCTCAACTTATATTTGGCTCAAATACCTTTCACTCT 933
Db	180 GlyserGlyValThrAsnGlyIleAlaAsnLeuVal---SerSerLysTyrrLeuLysAla 198
QY	934 GTTCAAGAAATACTATCTCATTTTCGCCGCATCTCGCTCGATTATTTCATCTCGAGGAACC 993
Db	199 AlaGlnGluLeuLeuaspGluValvalAsnalaspSerasp----- 212
QY	994 GAGTCAGGAGCTGCTAGTTCAGCGCTTACCTTCAGTGTGAGATATAFA---ACTGAGTTT 1050



Db 41 GlnGlnGlnGlnSerSerValProLeuSerSerLysLysTyrMetLysAlaAla 60  
QY 937 CAAGAAATATCTATCTATTCGCGCATCTCGCTCGATTAATCACTCGAGAACCGAG 996  
Db 61 GlnGlnLeuLeuAspGlu-----ValValAsnValGlyLysSerMetLysSerThrAsn 78  
QY 997 TCA---GGAGTGTAGTCTGAGCTTTTACCTTCAGGTTTGGAGATATACTAGTTT--- 1050  
Db 79 SerThrAspValValAsnAsnAspValLysSerLysAsnMetGlyAspMetAsp 98  
QY 1051 -----CTTGATGGT-----GATTCCTAATAACTCGGAGCGGGTTTCGGATCTACA 1095  
Db 99 GlyGlnLeuAspGlyValGlyAlaAspLysAspGlyAlaProThrThrGluLeuSerThr 118  
QY 1096 TTTCAAGGAGAGCATTAGAGCAAGCAAGAACCCATCTCTTGATCTTCTCAATGCGTG 1155  
Db 119 GlyGluArgGlnGluLeuLeuLysLeuAlaLysLeuValAsnMetLeuAspGluVal 138  
QY 1156 GATGATCGATATAGTCTGCTAGATGAGATCATACGGTTATATCAGGGTTCCATGCT 1215  
Db 139 GluGlnArgTyrArgHisTyrHisHisGlnMetGlnSerValIleHisTrpLeuGluGln 158  
QY 1216 GCACACGAGTAA---GATCCACAGTTACACCCCGTTTGCCTCCAAACCGTTCTTC 1272  
Db 159 AlalaGlyIleGlySerAlaLysThrTyrThrAlaLeuAlaLeuGlnThrIleSerLys 178  
QY 1273 TTATACAGAACTGAGAGAGAGAACTCGCAAGAGATA----- 1311  
Db 179 GlnPheArgCysLeuLysAspAlaIleIleGlyGlnIleArgSerAlaSerGlnThrLeu 198  
QY 1312 -----ATCTCTATGGATCTGATTGGAGAGAGCAAGAGCAAGAGCTCAAGAAACC 1362  
Db 199 GlyGluAspSerLeuGlyGlyLysIleGlyLysArgLeuLysPheValAspAsn 218  
QY 1363 TCTATGTTCCACAGCATTCCTCTCTCAGCAGCTCAAGAGCAAGAGCAAGATTCG 1422  
Db 219 GlnLeuArgGlnGlnArgAla---LeuGlnGlnLeuGlyMetIleGlnHisAsnLatrp 237  
QY 1423 AGACCTCAACGAGTTTGGCTGAGAAATCTGTTTCGGTTCTACGGAATGGATGTCCAA 1482  
Db 238 ArgProGlnArgGlyLeuProGluArgAlaValSerValLeuArgAlaTrpLeuPheGlu 257  
QY 1483 AACTTCCTTACCCCTTACCGAAAGATTGGAGAAACATCTCTAGCTATAGAGTGGC 1542  
Db 258 HisPheLeuHisProTyrProLysAspSerAspLysMetMetLeuAlaLysGlnThrGly 277  
QY 1543 TTGACAAAGAGTCAGGTATCAAACTGTTTATAATGCGCGGTTAGGCTATGGAAGCG 1602  
Db 278 LeuThrArgSerGlnValSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysPro 297  
QY 1603 ATGATAGAGAGATGATCGGAAATGAAACAAGAGGAGCTCAATAAC 1650  
Db 298 MetValGluGluMetTyrLeuGluGluIleLysGluHisGluGlnAsn 313

RESULT 9

QY028 PRELIMINARY; PRT; 393 AA.  
AC QY028; (TEMBLrel. 25, Created)  
DT 01-OCT-2003 (TEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
DE Bell-like homeodomain protein 1 (Fragment).  
GN BL1  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
OC lamids; Solanales; Solanaceae; Solanum.  
OX NCBI TaxID=4081;  
RN [1]\_"  
RP SEQUENCE FROM N.A.  
RA Ron N.; Parnis A.; Lifschitz E.;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AF375965; AAP47024.1; --  
KW DNA-binding; Homeobox; Nuclear protein.  
FT NON\_TER 1  
SQ SEQUENCE 393 AA; 44877 MW; D4A45EADA37F084 CRC64;  
Alignment Scores:  
Pred. No.: 6,38e-31 Length: 393  
Score: 450.50 Matches: 132  
Percent Similarity: 45.87% Conservative: 57  
Best Local Similarity: 32.04% Mismatches: 152  
Query Match: 13.72% Indels: 71  
DB: 10 Gaps: 13  
US-09-423-575-1 (1-1886) x QY028 (1-393)

QY 469 TTGTATATTTTCAATTTCTATCTGACTAACAATACTTATCATCTTCTCCACGGTCTATC 528  
Db 3 PheAspMetTyrGlnSerAspThrThrAlaTyrGlnProHisGlyGlyLeuSer--- 21  
QY 529 GATGTTCAAGATAACCGCAATGTTGAGTTTCATGCTCTCTCTCTCATCTCTCCCTCCACTT 588  
Db 22 -----ArgSerIleGluPheValAsnHisProAspPheThrThrAspSer 36  
QY 589 CATCTTTGGATCAATTTAAGACACTATGATGATTCCTCAAAACAACATGCGGGTTTGA 648  
Db 37 HisAspValAsnHisSerArgHisLeuMetAspLeuLeuGlyAlaSerHisAspAlaAsn 56  
QY 649 GCAATAGTGTGTTTCAGGCATTTTCAGGTGATGTTGTCAGTGAACCAATGATGCT 708  
Db 57 ThrAsnGlnGlnAlaGlnArgLeuSerLeuSerLeuGly---SerHisSerLeuValSer 75  
QY 709 ACATTTCGGTGAAGAAGATTTCCTGTTCTTAATTTTCG-----AATAAAGAAACAAT 759  
Db 76 ThrPheThrAsnAsn-----ProSerTyrMetAsnGlnGluIleAspGlnArgAsnAsn 93  
QY 760 GAGCTTTCAATGAGTCTTTCATCAGATGTTTCTGATGAATCTCGAGATAAGTCTTTGT 819  
Db 94 GluPheSerPheSer----- 98  
QY 820 GCAGCTACAGATGATGCTCAGCAAGCTTCTTCGACAGCAAGACATTTCTAATAAC 879  
Db 99 -----AlaAlaAlaMetAsnGlnSerPheSerAsnVal 109  
QY 880 GTTGTACTCAAGTTTCTCTCAACTATATTTGGCTCAAAATACCTCTCTGTTCAA 939  
Db 110 CysGlyThrGluSerPheValSerAlaIleGlyAsnSerLysTyrLeuLysProThrGln 129  
QY 940 GAAATACTATCTCATTTTCGCGCGATAC-----TCGCTCGATTATTC 981  
Db 130 SerLeuLeuGluGluLeuValCysIleGlyLysThrIleAspSerSerAsnGluLys 149  
QY 982 -----TCTCGAGGAACCGAGTCTGAGGCTGCTAGTTCTCAGCCTTTACTTCA 1026  
Db 150 PheIleArgArgLeuSerArgAsnSerLysGlySerLeuSer----- 164  
QY 1027 CGTTTTGAGAATACTAGTGTTCGTTGATTTCTAATACTCGGAGCGGGTTTC 1086  
Db 165 -----LeuArgAlaMetLeuLysGlyGluIleProProAsnAsnGluLeuPhe 180  
QY 1087 GGATCTCATTTCAAGAGGAGCATTAGAAGCAAGAAACCCATCTCTTGGATCTTCTT 1146  
Db 181 Asn-----GluArgHisGluLeuTyrValLysIleMetLysLeuIleAlaLeuLeu 197  
QY 1147 CAATGTGTGATGATCATATAGTTCATTCGTAGATGAGATTCATACGGTTATATCAGCG 1206  
Db 198 GluGluValGluArgArgTyrGluGlnTyrTyrGlnHisMetGluGluValThrSerThr 217  
QY 1207 TTCCATCTCTCAACCGGTTAGATCCA---CAGTTACACACCGGTTTGCCTCCAAACC 1263  
Db 218 PheGluValIleAlaGlyPheGlyAlaGlyAlaTyrThrAlaLeuAlaLeuGlnAla 237  
QY 1264 GTTTCCTCTTATACAAGAACCTGAGAGAGAGATCTGCAAGAGATAATCTCTATGGGA 1323  
Db 1323

```

Db 238 MetSerArgHisPheCysCysLeuArgAspSerIlelleSerGlnIleAsnPhelIleArg 257
QY 1324 TCTGTATTGGAGAGGAGGCAAGACAAAG-----ACTCAAGAAACCTCTATGTTTC 1371
Db 258 GlnLysMetProArgAspValProLysIleSerSerGlyLeuSerHisLeuSerLeuPhe 277
QY 1372 CACGACGATTCGCTT-----CTTCAGGAGCTGAAA-----CGAAGAAC 1410
Db 278 GluLysGluThrLeuGlnAsnArgIleSerLeuGlnGlnLeuGlyIleIleGlnSerAsn 297
QY 1411 CATCATGTTGGAGACCTCAACGAGGTTTGCCTGAGAAATCTGTTTCGGTTCTACGGAAT 1470
Db 298 ArgGlnAlaTrpGlnProIleArgGlyLeuProGluThrSerValAlaPheLeuArgSer 317
QY 1471 TGGATGTTCCAAACTTCCTTCAACCTTACCCGAAAGATTTCGGAGAAACATCTTCTAGCT 1530
Db 318 TrpLeuPheGluHisPheLeuHisProTyrProAsnAspSerGluLysLeuMetLeuSer 337
QY 1531 ATACGAAGTGGCTTGACAGAAGTCAGGTATCAAACTGGTTTATATAATCGCGGGTTAGG 1590
Db 338 SerGlnThrGlyLeuSerLysAsnGlnValSerAsnTrpPheIleAsnAlaArgValArg 357
QY 1591 CTATGGAACCGCATGATAGAAAGATGTATGCGGAA 1626
Db 358 LeuTrpLysProMetIleGluMetTyrLysGlu 369

```

## RESULT 10

```

Q8LLE1 PRELIMINARY; PRT; 532 AA.
AC Q8LLE1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE BEL1-related homeotic protein 14 (Fragment).
GN BEL14.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Desiree;
RA Chen H., Rosin F.M., Hannapel D.J.;
RT "A KNOX protein of potato interacts with several members of the TALE
RT family of transcription factors.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP046700; AN03624.1; -
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR006563; POX.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00574; POX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
FT NON TER 1
SQ SEQUENCE 532 AA; 59757 MW; 54940BF29B18474 CRC64;

```

```

Alignment Scores:
Pred. No.: 6.89e-31 Length: 532
Score: 450.50 Matches: 132
Percent Similarity: 46.51% Conservative: 61
Best Local Similarity: 31.81% Mismatches: 123
Query Match: 13.72% Indels: 99
DB: 10 Gaps: 17

```

US-09-423-575-1 (1-1886) x Q8LLE1 (1-532)

QY 610 CACTATGATGATTCCTCAAAACAACATGTGGGGTTTTTGAAGCAAAATAGTGAGTTTCAGGCA 669

```

Db 31 HisHisAspAspHisGlnGlySer---TrpHisHisAspAsnAsnArgThrLeuLeuVal 49
QY 670 TTTTCAGGTGTAGTTGGTCCAAGTGAACCAATGATGCTACATTCCGTGCAAGAGATTTC 729
Db 50 -----AspAspProSerMetArgCysVal-----Phe 58
QY 730 CCGTTTCTAATTTTGAATAAAGAAATAGCTTTTCATTGAGTCTTGCATCAGATGTT 789
Db 59 ProCysGluGlyAsnGlnArgProSerHisGlyLeuSerLeuSerLeuCys----- 75
QY 790 TCTGATCAATGCTCGAGATAAGTCTTTGTGAGCTTACAGATTAGCCTCAGACGAGCT 849
Db 76 SerSerAsnProSerSerIleGlyLeu-----GlnSerPhe 87
QY 850 TCTTGCAGCAGCAACACACATTTCTAATAAGTGTGTTTACTCAAGGTTTC----- 897
Db 88 GluLeuArgHisGlnAspLeuGlnGlnGlyLeuIleHisAspGlyPheLeuGlyLysSer 107
QY 898 -----TCTCAACTTATATTGCGCTCAAAATACCTT 927
Db 108 ThrAsnIleGlnGlnGlyTyrPheHisHisGlnValArgAspSerLysTyrLeu 127
QY 928 CACTCTGTTCAAGAAATACTATCTCATTTCCGCGCATACTCGCTC-----GAT 975
Db 128 GlyProAlaGlnGluLeuSerGluPheCysSerLeuGlyIleLysLysAsnAsp 147
QY 976 TATTCTATCT-----CGAGAACCGAGCTCAGAGCTCTAGTTTCAGCCTTT 1020
Db 148 HisSerSerLysValLeuLeuLysGlnHisGluSerThrAlaSerThrSerLys--- 166
QY 1021 ACTTCAGTTTTCAGAAATATACTAGTTCATTCGTTGATTCGTTTAACTCGGAGCG 1080
Db 167 -----LysGlnLeuLeuGlnSerLeuAsp----- 174
QY 1081 GGTTTCGATCTACATTTCAAAGGAGCATTAGAAGCAAGAAAGAAACCCATCTCTTGGAT 1140
Db 175 -----LeuLeuGluLeuGlnLysArgLysThrLysLeuLeuGln 187
QY 1141 CTCTCTCAATGTTGATGATGATATAGTATGATTCATTCGTTAGATGATTCATCGGTTATA 1200
Db 188 MetLeuGluGluValAspArgArgTyrLysHisTyrCysAspGlnMetLysAlaVal 207
QY 1201 TCAGCGTTTCCATGCT---GCAACCGAGTTAGATCCACAGTTACACACCGGTTGCGCTC 1257
Db 208 SerSerPheGluAlaValAlaGlyAsnGlyAlaAlaThrValThrSerAlaLeuAlaSer 227
QY 1258 CAAACCGTTTCTCTTCTTATACAAAGAACCTGAGAGAGAGAAATCTGCAAGAGATA----- 1311
Db 228 ArgAlaMetSerArgHisPheArgCysLeuArgAspGlyIleValAlaGlnIleLysAla 247
QY 1312 -----ATCTCTATGGATCTGTATTGGAGAGGAGGAAAGACAGACTCAA----- 1356
Db 248 ThrLysMetAlaMetGlyGlu-----LysAspSerThrSerThrLeuIle 262
QY 1357 -----GAAACG-----TCTATGTTTCCACGAG 1377
Db 263 ProGlySerThrArgGlyGluThrProArgLeuArgLeuLeuAspGlnThrLeuArgGln 282
QY 1378 CATTGCTTCTTTCAGCAGCTGAAACGAAAGAACCATCAGATTTGGAGACCTCAACGAGGT 1437
Db 283 GlnLysAlaPheGlnGlnMetAsnMetMetGluThrHisProTrpArgProGlnArgGly 302
QY 1438 TTGCTCGAGAAATCTGTTTCGGTTCTACGGAATTTGATGTTTCCAAACTTCCTTCACCTC 1497
Db 303 LeuProGluArgSerValSerValLeuArgAlaTrpLeuPheGluHisPheLeuHisPro 322
QY 1498 TACCCGAAAGATTCGGAGAAACATCTCTTACGTATACGAAGTGGCTGCACAAAGCTCAG 1557
Db 323 TyrProSerAspValAspLysHisIleLeuAlaArgGlnThrGlyLeuSerArgSerGln 342
QY 1558 GTATCAACCTGGTTTATAAATGCGCGGTTAGGCTATGGAAGCGGATGATAGAGAGATG 1617

```

Db 343 ValSerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetValGluGluMet 362  
QY 1618 TATCGGGAATGACCAAGAGAGAGCTCAATAACAGTCCACATTCACCAACGACCAACT 1677  
Db 363 TyrLeuGluThrLysGluGluAsnValGly-----SerProAspGly----- 378  
QY 1678 CTTGCAATGCCAAATCTGTATGATGAGCCCAAGCAATGATGATAA 1722  
Db 379 -----SerLysAlaLeuIleAspMetThrIleHisGln 390  
RESULT 11  
Q9FWS9 PRELIMINARY; PRT; 524 AA.  
AC Q9FWS9; 2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE FIB16.6 protein (BELL1-like homeodomain 3) (Homeodomain protein BELL1, putative).  
GN FIB16.6 OR BLH3  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,  
RA Altari H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,  
RA Buehler E., Chao Q., Chin C., Chioi J., Choi E., Gonzalez A.,  
RA Howing B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,  
RA Lenz C., Liu A., Liu S., Mukharbek N., Pham P., Sakano H., Shinn P.,  
RA Toriumi M., Vaynsberg M., Yu G., Ecker J., Theologis A., Davis R.W.;  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Pickowich M.S., Samach A., Modrusan Z., Haughn G.W.;  
RA "A family of BELL1-like homeodomain (BLH) proteins in Arabidopsis thaliana";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N., Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
RA "Full-length messenger RNA sequences greatly improve genome annotation";  
RL Genome Biol. 0:0-0(2002).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R., Feldmann K.;  
RA "Full-length cDNA from Arabidopsis thaliana";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC023754; ARG13065.1; -;  
DR EMBL; AF353093; AAK43835.1; -;  
DR EMBL; AY085278; AAW62510.1; -;  
DR FIR; F96784; F36784.  
DR HSSP; P41778; 1DU6.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003700; F:transcription factor activity; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 1.  
DR ProDom; PD000010; Homeobox; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
KW DNA-binding; Homeobox; Nuclear protein.  
SQ SEQUENCE 524 AA; 59679 MW; 942461E20FB3F39 CRC64;

Alignment Scores:  
Pred. No.: 2,58e-30 Length: 524  
Score: 444.00 Matches: 163  
Percent Similarity: 42.55% Conservative: 74

Best Local Similarity: 29.26% Mismatches: 170  
Query Match: 13.52% Indels: 150  
DB: 10 Gaps: 21  
US-09-423-575-1 (1-1886) x Q9FWS9 (1-524)  
QY 303 AATGGACAAACAACAACAACACCTTTAGTTCTCTGGAT----- 345  
Db 22 AsnGluGlnGlnGlnGlnGln-AlaserSerSerSerAlaAlaserPheSerG1 41  
QY 346 -----AATGTCATGACTAACAACAATCCTCTCTCATGTGATTTATACCTTC 392  
Db 41 uileValSerGlyAspValArgAsnAsnGlu-----MetValPheIleProPr 57  
QY 393 AAGAGAAGATTCAACTTCTCAATCTCAACAATGCTTCCATGGAATACCATCAGATCAGATCC 452  
Db 57 OthrSerAspValAlaValAsnGlyAsnVal-----ThrValSerSerAsnAs 73  
QY 453 TCTACAAATG-----GGTGGCTTTGATATTTTCAATTCTATGCTGACTACAAATACTT 506  
Db 73 PleuSerPheHisGlyGlyGlyLeuSerLeuSerLeuGlyAsnGlnIleGlnSerAlaVa 93  
QY 507 ATCATCTTCTCCACGGTCTATCGATGTTCAAGATAACCGCAATGTTGAGTTCATGGCTCC 566  
Db 93 lserValSerProPheGlnTyr----- 100  
QY 567 TCCTCTCATCTCTCTCCACTTCTCCTTTGGATCATTTAAGACACACTATGATCTCTC 626  
Db 101 -----HisTyrGlnAsnLeuSe 106  
QY 627 AAACACATCTGGGGTTTGAAGCAATAGTAGAOTTTTCAGGCATTTTCAGGTAGTTGG 686  
Db 106 rAsnGlnLeuSerTyrAsnAsnLeuAsnPro----- 116  
QY 687 TCCAAGTGACCAATGATCTACATTCGGTGAAGAAGATTTCCTCGTTCTTAATTCGAA 746  
Db 117 -----SerThrMetSerAspGlu----- 122  
QY 747 TAAAGAAACAATGAGCTTTCATGCTTCATGATGATTTCTGATGAATCTCGGA 806  
Db 123 -----AsnGlyLysSerLeuSerValHisGlnHisSerAspGlnIleLeuPr 139  
QY 807 GATAAGTCTTTGTGCAGCTACAAGATTAGCTCAGACCAAGCTCTTCGCACCAACAAGA 866  
Db 139 oSerSerVal-----TyrAsnAs 145  
QY 867 CATTTCTAATAACGTTGTT-----ACTCAAGGTTTCTC 899  
Db 145 nAsnGlyAsnAsnGlyValGlyPheTyrAsnAsnTyrArgTyrGluThrSerGlyPheVa 165  
QY 900 TCAACTTATATTGGCTCAAAATACCTTCACTCTGTTCAAGAATACTATCTCATTTCCG 959  
Db 165 lSerSerValLeuArgSerArgTyrLeuLysProThrGlnGlnLeuLeuAspGluValVa 185  
QY 960 CGCATACTCGCTCGATTAT-----TCATCTCGAGGAACCGA 995  
Db 185 lSerValArgLysAspLeuLysLeuGlyAsnLysLysMetLysAsnAspLysGlyGlnAs 205  
QY 996 GTCAGGAGCTGATGTCAGCCTTTTACTTCAGGTTTGGAGATAATAACTGAGTTCTTGA 1055  
Db 205 pPheHisAsnGlySerSer-----AspAsnIleThrGlu-----As 217  
QY 1056 TGGTGATTCTAATAACCGGCGGGTTCGGATCTACATTTCAAGAGGAGAGCATTAGA 1115  
Db 217 pAspLysSerGlnSerGlnGlnLeu-----SerProSerGluArgGlnGlnLeuG1 234  
QY 1116 AGCAAGAAACCACTCTCTTGGATCTCTTCAAAATGGTGGATGATGATATATGATCATTTG 1175  
Db 234 nSerLysLysLysLeuLeuThrMetValAspGluValAspLysArgTyrAsnGlnTy 254  
QY 1176 CGTAGATGAGATTCATACGGTTTATACAGCGTTTCCATGCTGTCACCGAGTTAGAT---CC 1232  
Db 254 rHisHisGlnMetGluAlaLeuAlaSerPheGluMetValThrGlyLeuGlyAlaAl 274



QY	1233	ACAGTTACACACCGGTTGGCCCTCCAAACCGTTTCTCTTATATACAAGAACCTGAGAGA	1292
Db	274	alysProTyrThrSerValalaleuAsnArgIleSerArgHisPheargCysLeuArgAs	294
QY	1293	GAGAATCTGCAAGAAGATA--ATCTCTATGGGATCTGTATTGGAGAGAGCGCAAGACAA	1349
Db	294	pAlaIleLysGluGlnIleGlnValIleArgGlyLysLeuGlyGluArgGluThrSerAs	314
QY	1350	GACTCAAGAAACCTCTATGTTCCACCAGCATTGCCCTCTTCACGAGCTGAAACGAAAGAA	1409
Db	314	pGluGlnGlyGluArgIleProArgLeuArgTyrLeuAspGlnArgLeuArgGlnArg	334
QY	1410	C-----CATCAG-----ATTGGAGACCTCAACGAGGTTTGCC	1442
Db	334	gAlaLeuHisGlnGlnLeuGlyMetValArgProAlaTrpArgProGlnArgGlyLeuPr	354
QY	1443	TGAGAAATCTGTTTCGGTTCTACGGAATTGGATGTTCCRAAACTCTCTTCACCTTACCC	1502
Db	354	oGluAsnSerValSerIleLeuArgAlaTrpLeuPheGluHisPheLeuHisProTyrPr	374
QY	1503	GAAGAATTCGGAGAAACATCTTTAGCTATACGAAGTGGCTTGACAAAGATCAGGTATC	1562
Db	374	oLysGluSerGluLysIleMetLeuSerLysGlnThrGlyLeuSerLysAsnGlnValAl	394
QY	1563	AAATCGGTTTATAATCGCGGGTTAGGCTATGGAAGCCCATGATAGAAGAGATGATGC	1622
Db	394	aAsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetIleGluGluMetTyrLy	414
QY	1623	GGAA-----ATGAACAAGGAAGAGCTCAATAACACGATTCACATTCACCCACGGACC	1673
Db	414	sGluGluPheGlyGluSerAlaGluLeuLeuSerAsnSer-----AsnGlnAs	430
QY	1674	AACTTCTCGAATGCCAAATCTGTTATGATGAGCGCAACGATGCATAA-----	1722
Db	430	pThrLysLysMetGlnGluThr-----SerGlnLeuLysHisGluAspSerSerSe	447
QY	1723	-----TAAGACAACAATGTGTTTACCACTTTGTGTGATAATTA	1760
Db	447	rSerGlnGlnGlnAsnGlnGlyAsnAsnAsnAsnIleProTyrThrSerAsp-Alag	467
QY	1761	GGCAATGTCTACTCTCATGTTGCCCAAAACCTAAACCATGTACGACT	1807
Db	467	luGlnAsnLeuValPheAlaAspProLysProAspArgAlaIleThr	482
RESULT 12			
Q8W412		PRELIMINARY; PRT; 524 AA.	
ID	Q8W412		
AC	Q8W412;		
DT	01-MAR-2002 (TrEMBLrel. 20, Created)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)		
DE	Similar to homeodomain proteins.		
GN	AT1G75410, F1B16.6 OR AT1G75410.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi		
NCBI	NCBI_TaxID=3702;		
EN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,		
RA	Palin C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,		
RA	Chen R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,		
RA	Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,		
RA	Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,		
RA	Ecker J., Theologis A., Davis R.W.;		
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,		
RA	Palin C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,		
RA	Chen R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,		

RA	Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA	Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinzaki K.,
RA	Ecker J., Theologis A., Davis R.W.;
RL	Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AY062545; AAL32623.1; -
DR	EMBL; BT000133; AN18452.1; --
DR	GO; GO:0005634; C:nucleus; IEA.
DR	GO; GO:0003700; F:transcription factor activity; IEA.
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR	InterPro; IPR001356; Homeobox.
DR	InterPro; IPR006563; FOX.
DR	Pfam; PF00046; homeobox; 1.
DR	P-domain; PD000010; Homeobox; 1.
DR	SMART; SK00389; HOX; 1.
DR	SMART; SM00574; FOX; 1.
DR	PROSITE; PS00027; HOMEBOX_1; 1.
DR	PROSITE; PS00071; HOMEBOX_2; 1.
KW	DNA-binding; Homeobox; Nuclear protein.
QZ	SEQUENCE 524 AA; 59693 MW; C9108DF9389A728E CRC64;

  

Alignment Scores:	
Pred. No.:	2,58e-30
Score:	Length: 524
Percent Similarity:	Matches: 163
Best Local Similarity:	Conservative: 74
Query Match:	Mismatches: 170
:	Indels: 150
:	Gaps: 21
DB:	

  

US-09-423-575-1 (1-1886) x Q8H4I2 (1-524)	
303 AATGGACAAACAACAACAACAACACTTTTATTGCTCTCGAT-----	345
22 ASnGLUGlnGlnGlnGlnGlnGln-AlaserSerSerAlaaserPheSerGI 41	
346 -----AATGTCACTACTAACCAAATCCTCTTCATGGATTTTATACCTTC 392	
:    :    :    :    :    :    :    :    :	
41 uLevelSerGlyAspValargasnanglu-----MetValPheileProPr 57	
393 AAGAGAAGATCAACTTCATCTCAACAATGCTCCATGGAATACATCATGATGCC 452	
:    :    :    :    :    :    :    :	
57 oTherSerAspLeuAlaValasnGlyasnVal-----ThrValSerSerAsnAs 73	
453 TCTACAATG----GGTGCGCTTGATATTTTCAATTCATGCTGACTAACAAATACTT 506	
73 pLeuSerPheHisGlyGlyLeuSerLeuGlyAsnGlnleGlnSerAlaVa 93	
507 ATCATCTCTCCACGGTCTATCGATGTTCAAGATAACCGCAATGTTGAGTTCATGGCTCC 566	
93 lserValserProphegIntyr----- 100	
567 TCCTCTCATCTCTCCCTCCATCTCATCTCTTGGATCATTTAAGACATCATGATGCCTC 626	
:    :    :    :    :    :    :    :    :	
101 -----HistyrlGlnasnLeuSe 106	
627 AAACAACATGTGGGTTTTGAAGCAATAGTAGTTTCAGGCATTTTCAGGTGTAGTTGG 686	
:	
106 rAsnGlnLeuSerTyrrAsnAsnLeuasnPro----- 116	
687 TCCAAGTGACCAATGATGTCTACATTCGGTGAAGAGATTCOCGTTTCTAATTCGAA 746	
:    :    :    :    :    :    :    :    :	
117 -----SerThrMetSerAspglu----- 122	
747 TAAAGAAAACAATGAGCTTTTCATGTCTTCATGATCATGATGTTTCTCATGAATCTCGA 806	
123 -----AsnGlylysSerLeuSerValHisGlnHisHisSerAspGlnlleLeuPr 139	
807 GATTAAGTCTTTGTGCAGCTACAAGATTAGCTTCAGACAGACCTTCTGCACGACCAAGA 866	
:    :    :    :    :    :    :    :    :	
139 oSerServal-----TyrAsnAs 145	
867 CATTTCTTAATAACGTTGTT-----ACTCAAGGTTTCTC 899	
145 nAsnGlyAsnAsnGlyValGlyPheTyrrAsnAsnTyrrArgtyrGluThrSerGlyPheVa 165	



QY 900 TCAACTATATATGGCTCAAAATACCTCTACTCTGTTCAGAAATACCTATCTCATTTCCG 959  
 Db .....  
 QY 165 lSerSerValLeuArgSerArgIyLeuLysProThrGlnGlnLeuLeuAspGluValVa 185  
 Db .....  
 QY 960 CGCATACTCTCGCTCATTTAT .....TCATCTCGAGAACCGA 995  
 Db .....  
 QY 185 lSerValArgLysAspLeuLysLeuGlyAsnLysLysMetLysAsnAspLysGlyGlnAs 205  
 Db .....  
 QY 996 GTGAGGAGCTGTAGTACCTTTACTTACCTTTTGTGAGATATACCTAGTCTTTTGA 1055  
 Db .....  
 QY 205 pPheHisAsnGlySerSer .....AspAsnIleThrGlu .....As 217  
 Db .....  
 QY 1056 TGTGTATTTCTAATAACTCGGAGCGGGTTCGGATCTACATTTCAAAGAGAGCATTAGA 1115  
 Db .....  
 QY 217 pAspLysSerGlnSerGlnGluLeu .....SerProSerGluArgGlnGluLeuGl 234  
 Db .....  
 QY 1116 AGCAAGAAACCATCTCTTGTGATCTTCTTCAATGTTGGATGATCGATATAGTCATGG 1175  
 Db .....  
 QY 234 nSerLysLysSerLysLeuLeuThrMetValAspGluValAspLysArgTyArgnGlnTy 254  
 Db .....  
 QY 1176 CGTAGATGAGATTCATACGGTTTATATCAGCGTTTCCATGCTGCAACCGAGTTAGAT ---CC 1232  
 Db .....  
 QY 254 rHisHisGlnMetGluLeuAlaLeuSerSerPheGluMetValThrGlyLeuGlyAlaAl 274  
 Db .....  
 QY 1233 ACAGTTACACCCGGTTTGCCTCCAAACCGTTTCTTCTATACAAAGCTGAGAGA 1292  
 Db .....  
 QY 274 aLysProTyThrSerValAlaLeuAsnArgIleSerArgHisPheArgCysLeuArgAs 294  
 Db .....  
 QY 1293 GAGAAATCTGCAAGAAGATA---ATCTCTATGGATCTGATTGGAGAGAGCGCAAGACAA 1349  
 Db .....  
 QY 294 pAlaIleLysGluGlnIleGlnValIleArgGlyLysLeuGlyGluArgGluThrSerAs 314  
 Db .....  
 QY 1350 GACTCAAGAAACCTATGTTTCCACCAGCATTCCTTCTTCCAGCAGCTCAAAACGAAAGAA 1409  
 Db .....  
 QY 314 pGluGlnGlyGluArgIleProArgLeuArgTyLeuAspGlnArgLeuArgGlnGlnAr 334  
 Db .....  
 QY 1410 C-----CATCAG-----ATTGGAGACCTCAACGAGTTTGC 1442  
 Db .....  
 QY 334 gAlaLeuHisGlnGlnLeuGlyMetValArgProAlaIleArgProGlnArgGlyLeuPr 354  
 Db .....  
 QY 1443 TGAGAAATCTGTTTCGGTTCTACGGAATGGATGTTCCAAACTTCTTCCACCTTACCC 1502  
 Db .....  
 QY 354 oGluAsnSerValSerIleLeuArgAlaIlePheLeuGluHisPheLeuHisProTyPr 374  
 Db .....  
 QY 1503 GAAAGATTGGAGAAACATCTTCTAGCTATACGAGTGGCTTCACAAAGAGTCAAGTATC 1562  
 Db .....  
 QY 374 cLysGluSerGluLysIleMetLeuSerLysGlnThrGlyLeuSerLysAsnGlnValAl 394  
 Db .....  
 QY 1563 AAATCTGTTTATAAATCGCGGTTTAGGCTATGGAGCGGATCATAGAGAGATGTATGC 1622  
 Db .....  
 QY 394 aSerIlePheIleAsnAlaArgValArgLeuIlePheLysProMetIleGluGluMetTyLy 414  
 Db .....  
 QY 1623 GGAA-----ATGAACAAGAGGAGCTCAATAACAGTCAATTCACCAACCGGACC 1673  
 Db .....  
 QY 414 sGluGluPheGlyGluSerAlaGluLeuLeuSerAsnSer-----AsnGlnAs 430  
 Db .....  
 QY 1674 ACTCTTCGATCCCAAAATCTGTTATGATGACCCGAGCAATGCATAAA----- 1722  
 Db .....  
 QY 430 pThrLysLysMetGlnGluThr-----SerGlnLeuLysHisGluAspSerSerSe 447  
 Db .....  
 QY 1723 -----TAAGACAACTTGTGTTTACCAACTTGTGATAATTA 1760  
 Db .....  
 QY 447 rSerGlnGlnAsnGlnGlyAsnAsnAsnAsnIleProTyThrSerAsp-AlaG 467  
 Db .....  
 QY 1761 GGCATTTGCTACTCTATGATGTCGCAAACTTAAACCATGTAGACT 1807  
 Db .....  
 QY 467 lucGlnAsnLeuValPheAlaAspProLysProAspArgAlaThrThr 482  
 Db .....

RESULT 13

Q9AYD9

ID Q9AYD9

AC Q9AYD9;

PRELIMINARY; PRT; 642 AA.

DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Putative homeodomain protein.  
 GN OSUNBA000323.5.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzae; Oryza.  
 OX NCBI\_taxid=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Buehl C.V., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,  
 RA Zismann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,  
 RA Bowman C.L., Craven B., Uterback T.R., Khalak H., Feldblum T.V.,  
 RA Quackenbush J., White C., Salzberg S.L., Fraser C.M.;  
 RT "Oryza sativa chromosome 3 BAC OSUNBA000323 genomic sequence."  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC079736; AAK00972.1; -.  
 DR HSSP; P41778; 1DU6.  
 DR Gramene; Q9AYD9; -.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR001356; Homeobox.  
 DR InterPro; IPR005663; POX.  
 DR Pfam; PF00046; homeobox; 1.  
 DR ProDom; PD00010; Homeobox; 1.  
 DR SMART; SM00389; HOX; 1.  
 DR SMART; SM00574; POX; 1.  
 DR PROSITE; PS00027; HOMEBOX 1; 1.  
 DR PROSITE; PS00071; HOMEBOX 2; 1.  
 KW DNA-binding; Homeobox; Nuclear protein.  
 SQ SEQUENCE 642 AA; 71056 MW; 2832B897B3371589 CRC64;

## Alignment Scores:

Pred. No.: 7,5e-30 Length: 642  
 Score: 439.00 Matches: 134  
 Percent Similarity: 44.30% Conservative: 68  
 Best Local Similarity: 29.39% Mismatches: 146  
 Query Match: 13.37% Indels: 108  
 DB: 10 Gaps: 14

US-09-423-575-1 (1-1886) x Q9AYD9 (1-642)

QY 532 GTTCAAGATACCCCAATGTTGAGTTTCATGCTCTCTCTCTCTCTCTCTCTCTCTCAT 591  
 Db .....PheMetGluLeuProGlyHisProThrAlaIle...  
 QY 592 CTTTGGATCATTTAAGACACTATGATTCCTCAACACACATGTGGGTTTGAAGCA 651  
 Db .....  
 QY 71 -----SerGlnAspSerSerArgGluProAsnMetValAlaSer 84  
 QY 652 AATAGTGAGTTTCAGGCATTTTCAGGTGTAGTTCCTCAAGTGAACCAATGATCTCT--- 708  
 Db .....GlyProAlaLysAspMetArgAsnGlu 100  
 QY 85 TyrMetAspGlnA:GSerPhe-----GlyProAlaLysAspMetArgAsnGlu 100  
 Db .....  
 QY 708 ----- 708  
 Db 101 MetLeuMetHisLeuMetAspGlyAlaHisAsnAlaGlyAlaAspLeuIleHisAsnAsp 120  
 QY 709 ACATTTCGGTGAAGAAGATTTCCTCGTTTCTAATTCGAATTAAGAAACAAT----- 759  
 Db .....  
 QY 121 ThrHisSerSerAlaGlnIleGluPheGlyLeuLeuAsnAsnHisAsnSerMetSerVal 140  
 QY 760 -----GAGCTTTTCATTCAGTCTTTCATCATCATGATGTTTCTGATGAATGC 801  
 Db .....  
 QY 141 AlaProAlaProGlyGlnGlyLeuSerLeuSerLeuAsnThrHisIleLeuAlaPro--- 159  
 QY 802 TCGGAGATAAGTCTTTGTGCGAGCTACAGATTACCTCAGAGCAAGCTTCTTCAGCAGC 861  
 Db .....  
 QY 160 SerTyProTyTrpSerAlaLysThrGluLeuLeuThrProHisSerTyHisGlyAsp 179  
 Db .....



```

QY 1357 -----GAAACC-----TCATGTTCCACGAGTTCCTTCTT 1389
Db 245 ArgGlyGluThrProArgLeuArgLeuLeuAspGlnThrLeuArgGlnGlnLysAlaPhe 264
QY 1390 CAGACGTTGAACGAAAGAACCATCATGATTTGAGACCTCAACGAGGTTTGCTGAGAAA 1449
Db 265 GlnGlnMetAsnMetMetGluThrHisProTyrArgProGlnArgGlyLeuProGluArg 284
QY 1450 TCTGTTTCGTTCTACGGAATTCGATGTTCCAAACTTCTTCAACCTTACCGGAAGAT 1509
Db 285 SerValSerValLeuArgAlaThrPheGluHisPheLeuHisProTyrProSerAsp 304
QY 1510 TCGAGAAACATCTTCTAGCTATACCAAGTGGCTTGACAAAGTCAAGTCAAGTATCAACTGG 1569
Db 305 ValAspLysHisIleLeuAlaArgGlnThrGlyLeuSerArgSerGlnValSerAsnTrp 324
QY 1570 TTTATAAATCGCGGTTAGGCTATGGAAGCCGATGATAGAGAGATGTATCGGAATG 1629
Db 325 PheIleAsnAlaArgValArgLeuTrpLysProMetValGluGluMetTyrLeuGluGlu 344
QY 1630 AACAGAGGAAG----- 1641
Db 345 ThrLysGluGluGluGluGluAsnValGlySerGlnAspGlySerLysAlaLeuLeuAsp 364
QY 1642 -----CTCAATAC-----AGTCACATTCAACCAACGACGCAACTCTTCA 1683
Db 365 GluMetThrIleAsnAsnHisGlnSerSerHisIleValGlnLysProAsnLeuValArg 384
QY 1684 ATCCCAAAATCTCTTATGATGACCAAGCAATGCAT-----AAATAAGACAAC 1731
Db 385 IleGluSerGluCysIleSerSerIleIleAsnHisProHisAspLysAsnAspGln 404
QY 1732 AAT 1734
Db 405 Asn 405

RESULT 15
Q94K35 PRELIMINARY; PRT; 759 AA.
ID Q94K35;
AC Q94K35;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Homeodomain protein JUBEL1.
GN JUBEL1.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Bonus;
RA Mueller J., Wang Y., Franzen R., Santi L., Salamini F., Rohde W.;
RT "Interactions between barley TALE homeodomain proteins suggest a role
RT for protein-protein associations in the regulation of Knox gene
RT function."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP334758; AK38645.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR005653; POX.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00574; POX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
SQ SEQUENCE 759 AA; 80092 MW; 982CE4B3F4EA2A0 ARC64;

```

Alignment Scores:

```

Pred. No.: 8,11e-29 Length: 759
Score: 427.50 Matches: 130
Percent Similarity: 44.09% Conservative: 75
Best Local Similarity: 27.96% Mismatches: 179
Query Match: 13.02% Indels: 81
DB: 10 Gaps: 13

US-09-423-575-1 (1-1886) x Q94K35 (1-759)
QY 445 TCAGATCCTCTACAAATGGGTGGCTTTGATATTTTCAATTCTATGCTGACTAACAAATAC 504
Db 194 SerSerProLysGlnGlnGlyGlyMet-----AlaGly 204
QY 505 TTATCATCTTCTCCACGGTCT-----ATCGATGTTCAAGATAACCGCAATGTT 552
Db 205 LeuAlaThrAspProAlaAlaAlaMetGlnLeuPheLeuMetAsnProGlnGlnGln 224
QY 553 GAGTTCATGGCTCTCTCCCTCATCTCTCCCTCATCTCTTGGATCATTTTAAGACAC 612
Db 225 GlnSerArgSerProThrSerProProSerAspAlaGlnSerAlaIleGlnHis 244
QY 613 -----TATGATGATCTCTCAACACATGTGGGTTTGAAGCAAT 654
Db 245 HisGluAlaPheGlnAlaTyrGlyAsnAlaAlaSerSerPheGlyGlyGlyAlaGly 264
QY 655 ---AGTCAGTTTTCAGGCATTTTCAGGTGTAGTTGGTCCAAAGT-----GAACCAATGATG 705
Db 265 ValValGluGlyGlnGlyLeuSerLeuSerLeuSerProSerLeuGlnLeuGluMet 284
QY 706 TCTACATCGGTGAAGAAGATTTCCTGTTCTAATTTTCGAATAAAGAAACAATGAGCTT 765
Db 285 AlalysGlnAlaGluGlu-----LeuArgValArgAspGlyValLeu 298
QY 766 TCATTGATCTTGCATCAGATGTTTCTGATGATGCTCGGAGATAAGTCTTTGCGAGCT 825
Db 299 TyrPheAsnArgGlnGlnGlnGlnGlnGlnGlnAlaAlaSerValGlnGlnLeu 318
QY 826 ACAAGATTAGCTCAGACCAAGCTTCTTGCGAGCAAGAACAGACATTTCTAATAACGTTGTT 885
Db 319 ProMetAlaLeuHisGlyGlnValGlySerMetGlyGlnGlnLeuHisValGlyTyrGly 338
QY 886 ACTCAAGGTTTCTCTCAACTATATTTGGTCCAAATACCTTCCATCTGTTTCAGAAATA 945
Db 339 ProAlaGlyValAlaGlyValLeuArgAsnSerLysTyrThrArgAlaAlaGlnGluLeu 358
QY 946 CTATCTCATTTCCGCCGATCTCGCTCGATTATTCATCTCGAGAACCGAGTCAGAGCT 1005
Db 359 LeuAspGluPheCysSerValGlyArgGlyGlnThrIleLysGlyGlyArgGlyGly 378
QY 1006 GCTAGTTTCAGCTTACTTACGTTTTCAGAAATAATAACTAGTTTCTTGATGGTATCT 1065
Db 379 SerSerSer-----AsnProAsnAlaSerLysGlyGlyProSer 391
QY 1066 AATAACTCGGAGGGGGTTCGGATCTACA----- 1095
Db 392 SerSerGlyAlaAlaGlnSerProSerSerAlaSerLysGluProProGlnLeuSerPro 411
QY 1096 TTTCAAAGGAGAGCATTAGAAGCAAGAAACCCATCTCTTGGATCTTCTTCAAAAGGTG 1155
Db 412 AlaAspArgPheGluGlnGlnArgLysAlaLysLeuIleSerMetLeuAspGluVal 431
QY 1156 GATGATCGATATAGTCTATTCGGTAGATGAGATTCATACGCTTATATCA----- 1203
Db 432 AspArgArgTyrAsnHisTyrCysAspGlnMetGlnMetValValAsnPheAspSer 451
QY 1204 -----OCGTTCCATGCTGCAACCGAGTTAGATCCACAGTTACACACCGGTTTGCCTC 1257
Db 452 ValMetGlyPheGlyAlaAlaThrPro-----TyrThrAlaLeuAlaGln 466
QY 1258 CAACCGGTTTCTCTTATACAGAAGCTGAGAGAGAGATCTGC----- 1302
Db 467 LysAlaMetSerArgHisPheArgCysLeuLysAspAlaIleAlaAlaGlnLeuArgHis 486

```

QY 1303 -----AAGAGATATCTCTATGGGATCTGTATTCGACAGAGCGAAA 1344  
Db ThrCysGluLeuGlyGluLysAspAlaGlyThrSerSerGlyLeuThrLysGlyGlu 506  
QY 1345 -----GACAAAGACTCAAGAAACCTCTATGTTCCACGACGAT 1380  
Db ThrProArgLeuArgAlaIleAspGlnSerLeuArgGlnArgAlaPheHisMet 526  
QY 1381 TCCCTTCTTACACAGCTGAACAAAGAACCATCATGTTGGAGACCTCAACGAGGTTTG 1440  
Db GlyMetMetGluGln-----GluAlaTrpArgProGlnArgGlyLeu 540  
QY 1441 CCTGAGAAATCGTTTCGGTTCTACGGAATTGGATGTTCCAAAACCTTCTTCCACCTTAC 1500  
Db ProGluArgSerValSerIleLeuArgSerTrpLeuPheGluHisPheLeuHisProTyr 560  
QY 1501 CCGAAAGATTCCGAGAAACATCTTCTAGCTATACGAGTGGCTTGACAAAGAGTCAGGTA 1560  
Db ProSerAspAlaAspLysHisLeuLeuAlaArgGlnThrGlyLeuSerArgAsnGlnVal 580  
QY 1561 TCAAACTGTTTATAAATCCGCGGTTAGGCTATCGAAGCCGATCATAGAGAGATGTAT 1620  
Db SerAsnTrpPheIleAsnAlaArgValArgLeuTrpLysProMetIleGluGluMetTyr 600  
QY 1621 GCGGAAATGACAAAGAGGAAGCTCAATAACAGTCACTTCAACCCACGACCAACTCTT 1680  
Db GlnGlnGluThrLysGluLeuGluGlySerSerAlaAlaAlaGlyGlyGlyVal 620  
QY 1681 CGAATGCCAAATCT 1695  
Db GlyGlyProGluSer 625

Search completed: September 2, 2004, 12:11:44  
Job time : 143 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2004, 05:19:27 ; Search time 4991 Seconds  
(without alignments)  
11284.326 Million cell updates/sec

Title: US-09-423-575-1  
Perfect score: 1886  
Sequence: 1 atttagttataaaattgttcg.....ttaaaaaaaaaaaaaaaaaa 1886

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vri:\*  
28: gb\_ges1:\*  
29: gb\_ges2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	508	26.9	511	9 AW004482	AW004482 701931784
C 2	411.4	21.8	423	9 AI999808	AI999808 701857547
C 3	406.2	21.5	423	9 AV808362	AV808362 AV808362
4	388.4	20.6	390	14 CB261894	CB261894 83-E8862-

5	277.6	14.7	773	14	CA917399	CA917399 EST641546
6	274.8	14.6	535	28	BH476676	BH476676 BQGN7677F
7	282.2	13.9	775	13	BQ990079	BQ990079 QGF15G17.
8	256.4	13.6	719	28	BZ033022	BZ033022 oen45H05.
C 9	251.6	13.3	270	9	AV800635	AV800635 AV800635
10	243	12.9	676	13	BU011811	BU011811 QGJ17D24.
C 11	242.4	11.8	709	28	BZ031456	BZ031456 cef98B09.
12	216.4	11.5	751	13	BU837920	BU837920 T107B12 P
13	147.4	7.8	672	13	BU024863	BU024863 QH6H21.Y
14	146	7.7	658	13	BU026692	BU026692 QH71713.
15	145.8	7.7	328	29	AL946255	AL946255 Arabidops
16	139.2	7.4	657	13	BU027235	BU027235 QHG3D05.Y
17	139.2	7.4	415	13	BQ848536	BQ848536 QGA7H07.Y
18	139.2	7.4	706	13	BQ859673	BQ859673 QGC13K23.
19	139.2	7.4	740	13	BQ867350	BQ867350 QGD10116.
20	135.2	7.2	777	14	CB343819	CB343819 CA32EN000
21	133.4	7.1	333	12	BG662137	BG662137 Ljirmpesc
22	133.4	7.1	669	13	BQ148285	BQ148285 NF065F11F
23	133.4	7.1	677	12	BG448606	BG448606 NF045H07N
24	133.2	7.1	360	13	BU765815	BU765815 sas20D01.
25	133	7.1	616	14	CA8319352	CA8319352 MCT026E10
26	132.8	7.0	850	14	CB292855	CB292855 UCRC801.0
27	132.6	7.0	393	9	AU294697	AU294697 AU294697_
28	132.4	7.0	706	12	B1272808	B1272808 NF092G09F
29	132	7.0	520	12	BM886468	BM886468 sam17a04.
30	132	7.0	555	12	BM520903	BM520903 eal31h09.
31	130.8	6.9	647	14	CD718786	CD718786 VVB154B04
32	130.8	6.9	658	14	CD717521	CD717521 VVB154B07
33	130.8	6.9	661	14	CD719099	CD719099 VVB157E07
34	130.8	6.9	675	14	CD718777	CD718777 VVB154A07
35	130.8	6.9	783	10	AW688195	AW688195 NF004E06S
36	130.2	6.9	445	10	AW596251	AW596251 s301B04.Y
37	130.2	6.9	479	9	AW102248	AW102248 sd85C07.Y
C 38	130	6.9	138	29	AL756855	AL756855 Arabidops
39	130	6.9	770	14	CA993653	CA993653 PAR09B07
40	129	6.8	570	14	CD056054	CD056054 H011A03S
41	128.8	6.8	537	12	BM554132	BM554132 sam70e12.
42	128.8	6.8	635	12	BM360047	BM360047 GA_sa002
C 43	128.6	6.8	673	28	BH556826	BH556826 BOHQJ62TF
44	128.6	6.8	480	13	BQ509372	BQ509372 EST616787
C 45	128.6	6.8	513	14	CD662087	CD662087 3529_1_13

#### ALIGNMENTS

RESULT 1  
AW004482/c 511 bp mRNA linear EST 08-SEP-1999  
LOCUS AW004482 701931784 A. thaliana, mixed source Arabidopsis thaliana cDNA clone  
DEFINITION AW004482 701931784, mRNA sequence.  
ACCESSION AW004482  
VERSION AW004482.1 GI:5851511  
KEYWORDS EST.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryote, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta, Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, rosids, eurosids II, Brassicales, Brassicaceae, Arabidopsis.  
REFERENCE 1 (bases 1 to 511)  
AUTHORS Chen, J., Momiya, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D., Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P., Gorgone, G., Burns, D., Griffin, J., Mouanoutou, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S., Nobrega, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and Hanson, D.  
TITLE Arabidopsis thaliana Gene Expression MicroArray  
JOURNAL Unpublished (1999)  
COMMENT Contact: David Smoller, Ph.D.  
Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc.  
4633 World Parkway Circle, St. Louis, MO 63134, USA

Tel: 877-577-2733  
Fax: 314-427-3324  
Email: service@genomesystems.com.

Carpio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S.,  
Nobriga, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and  
Hanson, D.  
Arabidopsis thaliana Gene Expression MicroArray  
Unpublished (1999)  
Contact: David Smoller, Ph.D.  
Genome Systems, Inc., a wholly owned subsidiary of Incyte  
Pharmaceuticals, Inc.  
4633 World Parkway Circle, St. Louis, MO 63134, USA  
Tel: 877-577-2733  
Fax: 314-427-3324  
Email: service@genomesystems.com.

## FEATURES

source  
1. .511  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/clone\_lib="A. thaliana, mixed source"  
/note="This sequence was obtained from a clone generated  
with a PCR product of the target gene."

## ORIGIN

Query Match 26.9%; Score 508; DB 9; Length 511;  
Best Local Similarity 99.4%; Pred. No. 2.8e-86;  
Matches 508; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1210 CATGCTGCAACCGAGTTAGATCCACAGTTACACACCCGGTTTGCCCTCCAAACCGTTTCC 1269  
Db 511 CATGCTGCANCGAGTTAGATCCACAGTTACACACCCGGTTTGCCCTCCAAACCGTTTCC 452  
QY 1270 TTCCTTATACAGAACTGAGAGAGAGAACTCTGCAAGAGAGAAATCTCTATGGATCTGTA 1329  
Db 451 TTCCTTATACAGAACTGAGAGAGAGAACTCTGCAAGAGAGAAATCTCTATGGATCTGTA 392  
QY 1330 TTGGAGAGAGGCAAGACCAAGAACTCTATGTTCCACGAGATTGCTTCTT 1389  
Db 391 TTGGAGAGAGGCAAGACCAAGAACTCTATGTTCCACGAGATTGCTTCTT 332  
QY 1390 CAGCAGCTGAACGAAAGAACCATCAGATTGGAGACCTCAACGAGGTTGCTTCCAGAAA 1449  
Db 331 CAGCAGCTGAACGAAAGAACCATCAGATTGGAGACCTCAACGAGGTTGCTTCCAGAAA 272  
QY 1450 TCTGTTTGGTCTACGAAATGGATGTTCCAAACCTTCTTCCCTTACCCGAAAGAT 1509  
Db 271 TCTGTTTGGTCTACGAAATGGATGTTCCAAACCTTCTTCCCTTACCCGAAAGAT 212  
QY 1510 TCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAGAGTCAGGTATCAAACTGG 1569  
Db 211 TCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAGAGTCAGGTATCAAACTGG 152  
QY 1570 TTTATAATGCGCGGTTAGCTATGGAAGCGGATGATAGAGAGATGATGCGGAATG 1629  
Db 151 TTTATAATGCGCGGTTAGCTATGGAAGCGGATGATAGAGAGATGATGCGGAATG 92  
QY 1630 AACAGAGAGAGCTCAATAACAGTCACATTCACCCACGACCAACTCTTCGAATGCCA 1689  
Db 91 AACAGAGAGAGCTCAATAACAGTCACATTCACCCACGACCAACTCTTCGAATGCCA 32  
QY 1690 AAATCTGTTATGATGAGCCAGCAATGCATA 1720  
Db 31 AAATCTGTTATGATGAGCCAGCAATGCATA 1

## RESULT 2

AI999808/c 423 bp mRNA linear EST 08-SEP-1999  
LOCUS 701557547 A. thaliana, Columbia Col-0, rosette-3 Arabidopsis  
DEFINITION thaliana cDNA clone 701557547, mRNA sequence.  
ACCESSION AI999808  
VERSION AI999808.1 GI:5846713  
KEYWORDS EST.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 423)  
Chen, J., Momiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D.,  
Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P.,  
Gorgone, G., Burns, D., Griffin, J., Mouantoutou, M., Nguyen, D.,  
Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C.,

## FEATURES

source  
1. .423  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/cultivar="Columbia Col-0"  
/db\_xref="taxon:3702"  
/clone\_lib="701557547"  
/tissue\_type="rosette"  
/dev\_stage="4 - 7 weeks"  
/clone\_lib="A. thaliana, Columbia Col-0, rosette-3"  
/note="Vector: pSPORT; Site 1: NotI; Site 2: SalI; cDNA  
library was derived from untreated rosette tissue from  
Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks.  
Plants were grown in 1:1:1 peat moss/vermiculite/perlite  
soil at 22 deg. C +/- 3 deg. C under constant light, and  
watered with fertilizer. cDNA synthesis was initiated  
using a NotI-oligo(dT) primer. Double-stranded cDNA was  
blunted, ligated to SalI adaptors, digested with NotI,  
size-selected, and cloned into the NotI and SalI sites of  
the pSPORT vector."

## ORIGIN

Query Match 21.8%; Score 411.4; DB 9; Length 423;  
Best Local Similarity 99.5%; Pred. No. 6.1e-68;  
Matches 412; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1400 AACGAAAGAACCATCAGATTGGAGACCTCAACGAGGTTGCTCGAGAAATCTGTTTCGG 1459  
Db 414 AACGAAAGAACCATCAGATTGGAGACCTCAACGAGGTTGCTCGAGAAATCTGTTTCGG 355  
QY 1460 TTCTAGGAAATGGATGTTCCAAACCTTCTTCCCTTACCCGAAAGATTCGGAGAAC 1519  
Db 354 TTCTAGGAAATGGATGTTCCAAACCTTCTTCCCTTACCCGAAAGATTCGGAGAAC 295  
QY 1520 ATCTTCTAGCTATACGAAGTGGCTTGACAAGAGTCAGGTATCAAACTGGTTTATAAATG 1579  
Db 294 ATCTTCTAGCTATACGAAGTGGCTTGACAAGAGTCAGGTATCAAACTGGTTTATAAATG 235  
QY 1580 CCGGGTTAGGCTATGGAAGCCGATGATAGAGAGATGATGCGGAATGAACAAGAGGA 1639  
Db 234 CCGGGTTAGGCTATGGAAGCCGATGATAGAGAGATGATGCGGAATGAACAAGAGGA 175  
QY 1640 AGCTCAATAACAGTCACATTCACCCACGACCAACTCTTCGAATGCCAAATCTGTGA 1699  
Db 174 AGCTCAATAACAGTCACATTCACCCACGACCAACTCTTCGAATGCCAAATCTGTGA 115  
QY 1700 TGATGAGCCAAAGATGCAATAAATGAACAACAATGTTGTTTACCAATTTGTGTAAT 1759  
Db 114 TGATGAGCCAAAGATGCAATAAATGAACAACAATGTTGTTTACCAATTTGTGTAAT 55  
QY 1760 AGGCAATGCTACTACTATGATGCGCCAAACCTTAACCATGTACCATCTATCAT 1813  
Db 54 AGGCAATGCTACTACTATGATGCGCCAAACCTTAACCATGTACCATCTATCAT 1

## RESULT 3

AV808362/c 423 bp mRNA linear EST 29-MAR-2002  
LOCUS AV808362 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-54-G14 3',  
DEFINITION mRNA sequence.  
ACCESSION AV808362

```

VERSION AV808362.1 GI:19842347
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE Arabidopsis thaliana; Embryophyta; Tracheophyta;
AUTHORS Eukaryota; Viridiplantae; Streptophyta; core eudicots;
Spermatophyta; Magnoliophyta; eudicotyledons; Brassicales; Arabidopsis.
1 (bases 1 to 423)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.,
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
2-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4339
Fax: 81-298-36-9060
Email: msek@rtc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified lambda FLN-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pluscript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.

FEATURES             Location/Qualifiers
     source            1..423
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /db_xref="taxon:3702"
                     /clone="RAF109-54-G14"
                     /dev_stage="plants at various developmental stages from
                     germination to mature seeds"
                     /lab_host="DH10B"
                     /clone_lib="PAP19"
                     /note="Site 1: BamHI; Site 2: SalI; subjected to
                     dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24
                     hr) treatments"

ORIGIN
Query Match          21.5%; Score 406.2; DB 9; Length 423;
Best Local Similarity 99.1%; Pred. No. 5.9e-67;
Matches 419; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1434 AGGTTTCCTGAGAAATCTCTTTCGGTTCTACGGAATTGGATTTCCAAACTTCCTTCA 1493
DB 423 AGGTTTCCTGAGAAATCTCTTTCGGTTCTACGGAATTGGATTTCCAAACTTCCTTCA 364
QY 1494 CCCTTACCCGAAGATTCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAGAG 1553
DB 363 CCCTTACCCGAAGATTCGGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAGAG 304
QY 1554 TCAGGTATCAAACTGGTTTAAATCGCGGGTAGGCTATGGAAGCCGATGATAGAAGA 1613
DB 303 TCAGGTATCAAACTGGTTTAAATCGCGGGTAGGCTATGGAAGCCGATGATAGAAGA 244
QY 1614 GATGATGCGGAATGAAACAGAGAGAGCTCAATACAGTCACATTCACCCCAACGACC 1673
DB 243 GATGATGCGGAATGAAACAGAGAGAGCTCAATACAGTCACATTCACCCCAACGACC 184
QY 1674 AACTCTTCGAATGCAAAATCTGTTATGATGAGCAGCAATGCATTAATAGACACAA 1733
DB 183 AACTCTTCGAATGCAAAATCTGTTATGATGAGCAGCAATGCATTAATAGACACAA 124
QY 1734 TTGTTTACCAACTTTGTGATTAATAGGCAATGTCTCTATGATGTCGCAAAACCTTA 1793
DB 123 TTGTTTACCAACTTTGTGATTAATAGGCAATGTCTCTATGATGTCGCAAAACCTTA 64
QY 1794 AACCATGACGACTATCAATACGATGTTATTA-TTGTATATACAACTCCTTTATCTTG 1852

```

```

DB 63 AACCATGTACGACTATCATTACGTATGTTATAAGTTGTATATACAACTCCTTTATCTTTG 4
QY 1853 ACT 1855
DB 3 ACT 1

RESULT 4
LOCUS CB261894
DEFINITION 83-E8862-008-010-K04-pal2 MP12 ADIS-008 Arabidopsis thaliana cDNA
clone MP12p767K0410Q 5-PRIME, mRNA sequence.
ACCESSION CB261894
VERSION CB261894.1 GI:32886667
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE Arabidopsis thaliana; Embryophyta; Tracheophyta;
AUTHORS Eukaryota; Viridiplantae; Streptophyta; core eudicots;
Spermatophyta; Magnoliophyta; eudicotyledons; Brassicales; Arabidopsis.
1 (bases 1 to 390)
Schmid, K.J., Sorensen, T.R., Stracke, R., Torjek, O., Altmann, T.,
Mitchell-Olds, T. and Weishaar, B.
Large-scale identification and analysis of genome-wide
single-nucleotide polymorphisms for mapping in Arabidopsis thaliana
Genome Res. 13 (6), 1250-1257 (2003)
JOURNAL 22883290
MEDLINE 12799357
PUBMED
COMMENT Contact: Weishaar B
ADIS DNA core facility at MP12
Max-Planck-Institute for Plant Breeding Research
Carl-von-Linne Weg 10, 50829 Koeln, Germany
Fax: 00492215062851
Email: weishaar@mpiz-koeln.mpg.de
Insert Length: 390 Std Error: 0.00
Plate: 10 row: K column: 04
Seq primer: pB12; GGTGGCGCCGCTCTAG.

FEATURES             Location/Qualifiers
     source            1..390
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /cultivar="C24"
                     /db_xref="GABI:601141"
                     /db_xref="taxon:3702"
                     /clone="MP12p767K0410Q"
                     /tissue_type="seedling"
                     /dev_stage="few days old seedlings"
                     /lab_host="E. coli XL1-Blue MRF"
                     /clone_lib="MP12-ADIS-008"
                     /note="Vector: pBluescript SK (-); Site 1: EcoRI; Site 2:
                     XhoI; cDNA library from Arabidopsis thaliana, accession
                     C24; seedling; Lambda ZAPII phage library was made at the
                     Max-Planck-Institute of Molecular Plant Physiology, Golem,
                     Germany and mass-excised at the Max-Planck-Institute for
                     Plant Breeding Research, Cologne, Germany; cloning sites
                     EcoRI-XhoI; Note: Sequencing granted in the context of the
                     'Establishment of high-efficiency SNP-based mapping tools
                     and development of methods for genome-wide mutation
                     detection' PI: Bernd Weishaar Sequence submission managed
                     by RZPD/GABI-Primary database: http://gabi.rzpd.de This
                     clone is available from RZPD; contact RZPD (clone@rzpd.de)
                     for further information."

ORIGIN
Query Match          20.6%; Score 388.4; DB 14; Length 390;
Best Local Similarity 99.7%; Pred. No. 1.4e-63;
Matches 389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1436 GTTTCCTGAGAAATCTCTTTCGGTTCTACGGAATGGATTTCCAAACTTCCTTCA 1495
DB 1 GTTTCCTGAGAAATCTCTTTCGGTTCTACGGAATGGATTTCCAAACTTCCTTCA 60

```

QY 1496 CTTACCCGAAAGATTCCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAGAGTC 1555  
 Db 61 CTTACCCGAAAGATTCCGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAGAGTC 120  
 QY 1556 AGGTATCAAACTGGTTTATAAATGCGCGGTTAGGCTATGGAAGCCGATGATAGAGAGA 1615  
 Db 121 AGGTATCAAACTGGTTTATAAATGCGCGGTTAGGCTATGGAAGCCGATGATAGAGAGA 180  
 QY 1616 TGTATGCGGAAATGAACAAGAGAAAGCTCAATAACAGTACACATTCACCAACCAAGACCAA 1675  
 Db 181 TGTATGCGGAAATGAACAAGAGAAAGCTCAATAACAGTACACATTCACCAACCAAGACCAA 240  
 QY 1676 CTCCTCGAATGCCAAATCTGTATGATGAGCCAAAGCAATGCAATAAAGACAACAAAT 1735  
 Db 241 CTCCTCGAATGCCAAATCTGTATGATGAGCCAAAGCAATGCAATAAAGACAACAAAT 300  
 QY 1736 GTGTTTACCAACTTGTGATAATTAGGCAATGCTACTCTATGATTGCCCAAAACCTAAA 1795  
 Db 301 GTGTTTACCAACTTGTGATAATTAGGCAATGCTACTCTATGATTGCCCAAAACCTAAA 360  
 QY 1796 CCATGACGACTATCATTTACGTTATTTATA 1825  
 Db 361 CCATGACGACTATCATTTACGTTATTTATA 390

## RESULT 5

LOCUS CA917399 773 bp mRNA linear EST 27-DEC-2002  
 DEFINITION EST641546 GPOD Medicago truncatula cDNA clone GPOD-36E1, mRNA  
 sequence.

ACCESSION CA917399.1 GI:27404329

## VERSION

KEYWORDS Medicago truncatula (barrel medic)

## SOURCE

Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.

REFERENCE 1 (bases 1 to 773)

AUTHORS Grusak, M.A., Samac, D.A., Town, C.D., Van Aken, S., Utterback, T.,

Cheung, F. and Fraser, C.M.

TITLE More ESTs from developing reproductive tissues of Medicago

JOURNAL truncatula

COMMENT Unpublished (2002)

Contact: Grusak, M.A.

USDA/ARS Children's Nutrition Research Center

Baylor College of Medicine

1100 Bates Street, Houston, TX 77030-2600, USA

Tel: 713 798 7044

Fax: 713 798 7078

Email: mgrusak@bcm.tmc.edu

TIGR sequence name: MTOB25TK

More information is available at: [www.medicago.org](http://www.medicago.org)

Seq primer: SKmod (CTA GAA CTA 9tg GAT CC).

## FEATURES

source

1. .773

/organism="Medicago truncatula"

/mol\_type="mRNA"

/cultivar="A17"

/db\_xref="taxon:3880"

/clone="GPOD-36E1"

/dev\_stage="Immature pod walls"

days after pollination"

/clone\_lib="GPOD"

/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:

XhoI; Immature pods, ranging in age from 15 to 30 days

after pollination, were collected from greenhouse-grown

plants. At harvest, seeds were removed from pods and

isolated pod walls were collected and immediately frozen

in liquid nitrogen. Pod walls were pooled for mRNA

extraction. cDNA was prepared from polyA+ enriched RNA.

The cDNA was directionally ligated into the Unizap XR  
 vector from Stratagene and packaged using Gigapack III  
 Gold packaging extracts. plasmids containing cDNA inserts  
 were excised from the recombinant lambda-Zap phage using  
 Ex-assist helper phage and propagated in XL04R cells."

## ORIGIN

Query Match 14.7%; Score 277.6; DB 14; Length 773;  
 Best Local Similarity 69.9%; Pred. No. 1.2e-42;  
 Matches 411; Conservative 0; Mismatches 159; Indels 18; Gaps 2;  
 QY 1068 TAACCTGGAGCGGGTTTCGGATCTACATTTCAAGAGAGCAATTAGAGCAAGAAAAC 1127  
 Db 6 TATGTCAGAAGCTTATGCTGATTCCTTCATGCAAGGCATGCTGTTGAATCAAGCAATC 65  
 QY 1128 CCATCTCTTGGATCTTCTTCAAAATGGTGGATGATCGATATAGTCATTTGGTATGAGAT 1187  
 Db 66 TCAACCTCTTGGATCTTCTTCAAAATGGTGGATGATCGATATAGTCATTTGGTATGAGAT 125  
 QY 1188 TCATAGCGTTATATCAGCGTTCCATGCTGCAACCGAGTTAGATCCACAGTTACACCCG 1247  
 Db 126 TCACACTGTGCTATCTGCAATTTTCATGCTGCTGAGCTCGATCCACAAATACACCGCA 185  
 QY 1248 GTTTGCGCTCCAAACCGTTTCTTCTTATACAAAGAACCTTGAGAGAGAGATCTGCAAGAA 1307  
 Db 186 CTTTGTCTGTCAAACCGTTTCTTCTTATAGGACTTGAAGAGAGAGATAGGAAACA 245  
 QY 1308 GATAATCTCTATGGGATCTGTATTTGGAGAG-----AGGCAAGACAGAAC 1352  
 Db 246 TATCTTTCCATGGGATCAAAATTTTAAACAGTTTCATGTCAGAAGAGAGATAAGGAATTATC 305  
 QY 1353 TCAGAAACCTCTATGTTCCACCGAGATTGCCTTCTTCAGCAGCTCAACGAGAGAACCA 1412  
 Db 306 TGTGAAACTTCATTCATTCAAAGCAATGGGCTCTTCAACAGTTGAA---AAGGAAAGA 362  
 QY 1413 TCAGATTGGAGACCTTCAACGAGGTTTGCTCGTGAAGAAATCTGTTTCGTTCTACGGAATTG 1472  
 Db 363 TCAGTTATGAGGCGCCCAAGGGGCTTGCAGAAAGATCTGTTTCGTTCTACGTTGATG 422  
 QY 1473 GATGTTCCAAAACCTCTCTACCCCTTACCCGAGAGATTCGAGAGAACATCTTCTAGCTAT 1532  
 Db 423 GATGTTTCAGAAATTTCTCCATCCGATCCGATCCTAAAGATGAGAGAGAGCATTTACTTCCAAT 482  
 QY 1533 ACAGAGTGGCTTGACAAGAGTCAAGTATCAAACTGGTTTATAAATGCGCGGTTAGGCT 1592  
 Db 483 AAAAGTGGTTGACTAGAGTCAGGTATCAAACTGGTTTATAAATGCGCGTTCGGCT 542  
 QY 1593 ATGGAAGCGGATGATAGAGAGATGTATCGGAAATGAACAAGAGAGAA 1640  
 Db 543 ATGGAACCACTTATTGAGGAAATGTATGCTGAAATGAATAGAGAGAA 590

## RESULT 6

BH476676

LOCUS

DEFINITION

BH476676

VERSION

KEYWORDS

SOURCE

ORGANISM

Brassica oleracea

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 535)

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2001)

Other\_GSSs: BGN787TR

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

BH476676 535 bp DNA linear GSS 13-DEC-2001  
 BGN787TR BGN Brassica oleracea genomic clone BGN787, genomic  
 survey sequence.

BH476676

GSS

Brassica oleracea

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 535)

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2001)

Other\_GSSs: BGN787TR

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.



for details.  
belongs to contig QG\_CA\_Contig4128, see <http://cgpdb.ucdavis.edu/>  
Email: akozikeatgc.org [michelmore@vegmail.ucdavis.edu]

```

location/quantity
source
1. .775
/organism="Lactuca sativa"

```

```

/mo1 type= "mura
/cultivar="L.serriola"
/db xref="taxon:4236"

```

```
/clone="QGF19G17"
```

```

/clone_lib="QG_EFGHU lettuce serriola"
/note="vector: pBRCNDAS1AB; The library was constructed
from 10 different sources of RNA from a single genotype.

```

Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-cap vector and

transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/mac/ciserm-fluoresce>.

```
ORIGIN
TAG LIB=QG EFGHJ lettuce serriola
TAG_SEQ=GCTTGACGGG"
```

Query Match 13.9%; Score 262.2; DB 13; Length 775;  
 Best Local Similarity 67.9%; Pred. No. 9.8e-40;  
 Matches 382; Conservative 0; Mismatches 178; Indels 3; Gaps 1;

Qy	1082	GTTTCGATCTACATTTCAAGGAGAGCATTTAGAGCAAGAAACCATCTCTTGGATC	1141
Dd	60	GATTTGTTCAATGCTTTTCAAAAGACGAAAGTGTCGAAGCTTAAAGAAAGCATTTGCTCGCCT	119

Qy	1142	TTCTTCAAATGGTGGATGATCGATATAGTCAATTCGCTAGATGAGATTTCATACGGTTATAT	1201
Db	120	TACTAGAAATGGTGGATGAGCGGTATAATCACTGTTTAGATGAGATACATACGGTGATAT	179

Qy	1202	CAGCGTTCCATGCTGCAACCGAGTTAGATCCACAGTTACACACCGGTTTGCCTCCAAA	1261
Dh	180	CGCGATTTCCAGCTGTGACACAGATTACATTCGGGAGGTCATCCCTGTTTGTCTTTCATA	239

[illegible]

Qy 1322 GATCTGTATT---GGAGAGAGGCAAGACAGACTCAAGAAACCTCTATGTTCCACGAGC 1378

Qy 1379 ATTGCCTTCTCAGCAGCTGAAAGAACCATCAGATTTCGAGACCTCAACGAGCTT 1438

QY 1439 TCCTGAGAACTCTGTTTCGGTCTACGGAATGGATGTTCCAAAACTTCCTTCACCCTT 1498

420 1ACCGGHHHHHCTGCGGAGAAACATCTTCTAGCTATACGAAAGTGGCTTGCACGAAGTCAGG 1558

DB	QY
480	ATCCAAAGGATCGGGAGAAACAATATATGGCTGTATAAAAAGCCGATATGACAAAGAGACCAGG
1559	TATCAAACTGGTTTATAAATCGCGGGGTAGGCTATCGAAGCCGATATAGAAGAGATGT

DB 540 TATCGAATTGGTTTATAAATCGAGGGTGAGACTATGGAAGCCCATGATTGAGGAAATGT 599

```

QY 1619 ATCGGAAATGAACAGAGGAG 1641
Db 600 ACTTAGAGATGAATAGGAGG 622

RESULT 8
BZ033022 719 bp DNA linear GSS 09-OCT-2002
LOCUS BZ033022 Brassica oleracea genomic, genomic survey
DEFINITION oeh45h05.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
sequence.
ACCESSION BZ033022
VERSION BZ033022.1 GI:23607215
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE Delaunay, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,
Nash, W., Rabinowicz, P.D. and Wilson, R.K.
Whole genome shotgun reads from Brassica oleracea
Unpublished (2002)
AUTHORS Contact: Richard K. Wilson
JOURNAL Genome Sequencing Center
COMMENT Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: oeh45 row: h column: 05
Seq primer: -21UpOT forward
Class: Shotgun
High quality sequence start: 16
High quality sequence stop: 551.
Location/Qualifiers
1. .719
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
/note="Vector: pOW13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea T01000DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center."

FEATURES
source
Query Match 13.6%; Score 256.4; DB 28; Length 719;
Best Local Similarity 66.3%; Pred. No. 1.3e-38;
Matches 456; Conservative 0; Mismatches 156; Indels 76; Gaps 3;

QY 789 TTCGATGATCTCGGAGATGAAGTCCTTTGTCAGCTACAGATTAGCCTCAGAGCAAGC 848
Db 31 TTCTGATGAGTCTCAGAGATGAAGTCCTTATGCTCTACTAAATGACCTCAGAGCAAGC 90

QY 849 TTCTTGCAGCAAGACATTTCTTAATAACGTTCTTACTCAAGTT---TCTCTCAACT 905
Db 91 TTTGACCTCAGAGCAAGCTTTCTCAGCAGCAAGACATTATATACGTTCTCTCATGT 150

QY 906 TATATTGGCTCAAAATACCTTCATCTGTGTCAGAAATATCTATCTATTTTCGGCGGATA 965
Db 151 TATATTGGCTCAAAAGTATCTTCACTCTGTTCAAGAAATACTATCTCAATTCGTACATA 210

QY 966 CTCGGCT-----CGATTATTCATCTCGAGGAACCGAGTC 998
Db 211 CTCCTCCACGCTTAGAGATAATCTCTCAATGCTATTTTTCATCTCAAGAAATCTAGTC 270

QY 999 AGSAGCTGTAGTTCAGCTTTTACTTCACGTTTGTGAGAAATATAACTGAGTTTCTTGATGG 1058
Db 271 AAGGGCTGTGGTTTCTAGCTTTTACTTCAGTTTACGTTAGAGAAATCAACCGGAGTTTCTTGAAG 330

QY 1059 TCAATTCTAATAA-----CT 1072

```

```

Db 331 AGTTTTTGATCAAAAGACGAGCATTAGAACGAAAGAAACCAGCTCTTGGATCTTCTTCA 390
QY 1073 CGGAGCGGGTTTCGGATCTACATTTTCAAGGAGAGCATTAGAACGAAAGAAACCCTC 1132
Db 391 AATGGTATATATTCACACATATATATATATATATATATATATATATATATATATAT 450
QY 1133 TCTTGATCTTCTTCAATGGTGATCGATATATATATATATATATATATATATATATAT 1192
Db 451 GAGTTGGATTTGCTCTCAGTGGATGAATAATATATATATATATATATATATATATAT 510
QY 1193 CGTTATATCAGGTTCCATGTCGTAACCGAGTTAGATCCAGTTTACACCCGGTTTG 1252
Db 511 CGGTTGATCAGCCTTCCATGCAGCAACCGAGTTAGACCCGCGAGCTACACCCGGTTTG 570
QY 1253 CCCTCAAAACCGTTTCTTCTTATACAAGAACTGAGAGAGAGAAATCTTCAAGAGATAA 1312
Db 571 CATCCAAACCATCTCTGTTCTTATACAGAACTGAGGAGAGATCAGCAAAAGATAC 630
QY 1313 TCTTATGGATCTGTATTGGAGAGAGGCAAGAACTCAAGAACTTATGTTCC 1372
Db 631 TTATGATGGATCAGTTTATAGAGAGAGGCAAGAGAGTCTCAAGAAATTCATCATCC 690
QY 1373 ACCAGATTGCCTTCTTCCAGCAGCTGAA 1400
Db 691 ACCAGATTGCCTTCTTCCAGCAGATGAA 718

RESULT 9
AV800635/c 270 bp mRNA linear EST 29-VAR-2002
LOCUS AV800635 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-24-021 3',
DEFINITION mRNA sequence.
ACCESSION AV800635
VERSION AV800635.1 GI:19834620
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 270)
AUTHORS Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
TITLE Plant Functional Genomics Research Group
JOURNAL RIKEN Genomic Sciences Center
COMMENT 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msekit@r.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified Bluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.
Location/Qualifiers
1. .270
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL09-24-021"
/dev_stage="plants at various developmental stages from
germination to mature seeds"
/lab_host="DH10B"
/clone_lib="RAFL9"
/note="Site 1: BamHI; Site 2: SalI; subjected to
dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24

```

hr) treatments"		source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <a href="http://cgdb.ucdavis.edu/">http://cgdb.ucdavis.edu/</a>	
TAG LIB=QG EFGHJ lettuce serriola		TAG_SEQ=GCTTGACGGG"	
ORIGIN	Query Match	12.9%; Score 243; DB 13; Length 676;	
	Best Local Similarity	69.9%; Pred. No. 4.5e-36;	
	Matches 343; Conservative	0; Mismatches 145; Indels	3; Gaps 1;
QY	1154	TGGATGATCGATATAGTCAATTCGCTAGATGAGATTCATACGGTTATATACGGTTTCATG	1213
Db	1	TGGATGATCGATATAGTCAATTCGCTAGATGAGATTCATACGGTTTCATG	60
QY	1214	CTGCAACCGAGTTAGATCCACAGTTACACCCCGTTTGCCCTCCAAACCGTTTCCTTCT	1273
Db	61	CTGTAAACAGAAATGAATCCGCAAGTTTCATGCTGTTTGTCTCTTCATACGATTACGTTT	120
QY	1274	TATACAGAACTCTGAGAGAGAGATCTGCAAGAGATAATCTCTATGGATCTGTATT--	1331
Db	121	TCTACAAAAGTTTAAGAGAGAGATTAAGTAATTATATCTTCAATGGGGCGGATTTA	180
QY	1332	-GGAGAGAGCAAGACCAAGACTCAAGAAACCTCTATGTTCCACAGATTCGCTTCTTC	1390
Db	181	GCACCATGATCAAGGAGAGATGAGTGTCTTGTACCAAGCAATGGGCTCTTC	240
QY	1391	AGCAGCTGAAACGAGAGACCATCAGATTTGGAGACCTCACGAGTTTGCTCGAGAAAT	1450
Db	241	AGCAGCTGAAAGAGAGATGATCAGCTATGGAGCGCGCAAGAGGGCTTCGGAAAT	300
QY	1451	CTGTTTCGGTTCTACGGAATTTGATGTTTCCAAAACCTTCCTTACCCCTTACCCGAAAGATT	1510
Db	301	CTGTTTCGGTTCTACGGAATTTGATGTTTCCAAAACCTTCCTTACCCCTTACCCGAAAGATT	360
QY	1511	CGGAGAAACATCTCTAGCTATAGAAAGTGCCTTGACAGAGTCAGTATCAACCTGTT	1570
Db	361	CGGAGAAACATCTCTAGCTATAGAAAGTGCCTTGACAGAGTCAGTATCAACCTGTT	420
QY	1571	TTATAAATGCCGGTTAGGCTATGGAGCCGATGATAGAGAGATGATCGGAAATGA	1630
Db	421	TTATAAATGCCGGTTAGGCTATGGAGCCGATGATAGAGAGATGATCGGAAATGA	480
QY	1631	ACAAGAGGAG 1641	
Db	481	ATAGAGGAGG 491	
RESULT 11	BZ031456/c	709 bp	DNA linear
LOCUS	BZ031456	709 bp	DNA linear
DEFINITION	cef98b09.g1 B.oleracea002 Brassica oleracea genomic, genomic survey		
ACCESSION	BZ031456		
VERSION	BZ031456.1	GI:23604956	
KEYWORDS	GSS.		
SOURCE	Brassica oleracea		
ORGANISM	Brassica oleracea		
REFERENCE	1 (bases 1 to 709)		
AUTHORS	Delehaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,		
TITLE	Nash, W., Rabinowicz, P.D. and Wilson, R.K.		
JOURNAL	Whole genome shotgun reads from Brassica oleracea		
COMMENT	Unpublished (2002)		
	Contact: Richard K. Wilson		
	Genome Sequencing Center		
	Washington University School of Medicine		
	Email: <a href="mailto:submissions@watson.wustl.edu">submissions@watson.wustl.edu</a>		

hr) treatments"		source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <a href="http://cgdb.ucdavis.edu/">http://cgdb.ucdavis.edu/</a>	
TAG LIB=QG EFGHJ lettuce serriola		TAG_SEQ=GCTTGACGGG"	
ORIGIN	Query Match	12.9%; Score 243; DB 13; Length 676;	
	Best Local Similarity	69.9%; Pred. No. 4.5e-36;	
	Matches 343; Conservative	0; Mismatches 145; Indels	3; Gaps 1;
QY	1154	TGGATGATCGATATAGTCAATTCGCTAGATGAGATTCATACGGTTATATACGGTTTCATG	1213
Db	1	TGGATGATCGATATAGTCAATTCGCTAGATGAGATTCATACGGTTTCATG	60
QY	1214	CTGCAACCGAGTTAGATCCACAGTTACACCCCGTTTGCCCTCCAAACCGTTTCCTTCT	1273
Db	61	CTGTAAACAGAAATGAATCCGCAAGTTTCATGCTGTTTGTCTCTTCATACGATTACGTTT	120
QY	1274	TATACAGAACTCTGAGAGAGAGATCTGCAAGAGATAATCTCTATGGATCTGTATT--	1331
Db	121	TCTACAAAAGTTTAAGAGAGAGATTAAGTAATTATATCTTCAATGGGGCGGATTTA	180
QY	1332	-GGAGAGAGCAAGACCAAGACTCAAGAAACCTCTATGTTCCACAGATTCGCTTCTTC	1390
Db	181	GCACCATGATCAAGGAGAGATGAGTGTCTTGTACCAAGCAATGGGCTCTTC	240
QY	1391	AGCAGCTGAAACGAGAGACCATCAGATTTGGAGACCTCACGAGTTTGCTCGAGAAAT	1450
Db	241	AGCAGCTGAAAGAGAGATGATCAGCTATGGAGCGCGCAAGAGGGCTTCGGAAAT	300
QY	1451	CTGTTTCGGTTCTACGGAATTTGATGTTTCCAAAACCTTCCTTACCCCTTACCCGAAAGATT	1510
Db	301	CTGTTTCGGTTCTACGGAATTTGATGTTTCCAAAACCTTCCTTACCCCTTACCCGAAAGATT	360
QY	1511	CGGAGAAACATCTCTAGCTATAGAAAGTGCCTTGACAGAGTCAGTATCAACCTGTT	1570
Db	361	CGGAGAAACATCTCTAGCTATAGAAAGTGCCTTGACAGAGTCAGTATCAACCTGTT	420
QY	1571	TTATAAATGCCGGTTAGGCTATGGAGCCGATGATAGAGAGATGATCGGAAATGA	1630
Db	421	TTATAAATGCCGGTTAGGCTATGGAGCCGATGATAGAGAGATGATCGGAAATGA	480
QY	1631	ACAAGAGGAG 1641	
Db	481	ATAGAGGAGG 491	
RESULT 11	BZ031456/c	709 bp	DNA linear
LOCUS	BZ031456	709 bp	DNA linear
DEFINITION	cef98b09.g1 B.oleracea002 Brassica oleracea genomic, genomic survey		
ACCESSION	BZ031456		
VERSION	BZ031456.1	GI:23604956	
KEYWORDS	GSS.		
SOURCE	Brassica oleracea		
ORGANISM	Brassica oleracea		
REFERENCE	1 (bases 1 to 709)		
AUTHORS	Delehaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,		
TITLE	Nash, W., Rabinowicz, P.D. and Wilson, R.K.		
JOURNAL	Whole genome shotgun reads from Brassica oleracea		
COMMENT	Unpublished (2002)		
	Contact: Richard K. Wilson		
	Genome Sequencing Center		
	Washington University School of Medicine		
	Email: <a href="mailto:submissions@watson.wustl.edu">submissions@watson.wustl.edu</a>		

Plate: oef98 row: b column: 09  
 Seq primer: -28RPpOT reverse  
 Class: shotgun  
 High quality sequence start: 16  
 High quality sequence stop: 551.  
 Location/Qualifiers  
 1..709  
 /organism="Brassica oleracea"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3712"  
 /clone\_lib="B.oleracea002"  
 /notes="Vector: pOTw13; Whole genome shotgun library from  
 flowering buds. DNA was purified from a crude nuclear  
 prep using Brassica oleracea TO1000DH3 buds provided by  
 Thomas Osborn at the University of Wisconsin. Genomic  
 DNA was provided by Pablo Rabinowicz (CSHL) and the  
 shotgun library prepared at Washington University Genome  
 Sequencing Center."

ORIGIN

Query Match 11.8%; Score 222.4; DB 28; Length 709;  
 Best Local Similarity 69.0%; Pred. No. 3.5e-32;  
 Matches 486; Conservative 0; Mismatches 86; Indels 132; Gaps 8;  
 Qy 496 AACATATCTTATCATCTCTCCACGGTCTATCGATGTTCAAGATAACCGCAATCTTGAG 555  
 Db |||||  
 Qy 708 AACATATCTTATCATCTCTCCACGGTCTATCGATGTTCAAGATAACCGCAATCTTGAG 652  
 Db |||||  
 Qy 556 TTATGGCT 615  
 Db |||||  
 Qy 651 GCCATGGCTCC-----TCT 601  
 Db |||||  
 Qy 616 GATGATCTCTCAACACATGTTGGGTTTGAAGCAATAGTGAATGAGTTTCAGGATTTTC- 674  
 Db |||||  
 Qy 600 GATGATCTCTCAACACATGTTGGGTTTGAAGCAATAGTGAATGAGTTTCAGGATTTTC- 541  
 Db |||||  
 Qy 675 -----AGGTGATGTTGGTCAAGTGAACCAATGATCTTACATTTGGTGAAGAAAGAT 726  
 Db |||||  
 Qy 540 TTATAGAGAGTTGTTGGTCTAGTGAACCAATATATCTTACATCTCGGTGAAGAA--- 484  
 Db |||||  
 Qy 727 TTCCCGTTCTAATTTTGAATAAAGAAACATGAGTTTTCATGAGTTTGCATTCAGAT 786  
 Db |||||  
 Qy 483 -----TTCCCAAGAAACGAGTTCTCGTTAAGTCTTGGCACTGAT 445  
 Db |||||  
 Qy 787 GTTTCGTGATGTCGCGAGATAGTCTTTGTGC-----AGCTACAGATGATGCC 837  
 Db |||||  
 Qy 444 GTTTCGTGATGTCGCGAGATAGTCTTTGTGCTGCTACTAATCTACTAGATAACC 385  
 Db |||||  
 Qy 838 TCAGACAAAGCTTTTTCAGACGACAAAGACATTTCTAATAACGTTGTTACTCAAGGTTTC 897  
 Db |||||  
 Qy 384 TCAGACAAAGCTTTTTCAGACGACAAAGACATTTCTAATAACGTTGTTACTCAAGGTTTC 343  
 Db |||||  
 Qy 898 TCTCAACTTATATTGGCTCAAAATACCTTCACTCTGTTTCAAGAAATATCTCATTTTC 957  
 Db |||||  
 Qy 342 TCTCAAGTCATATTGCTCCAGTACCTTCACTCTGTTTCAAGAAATATCTCATTTTC 283  
 Db |||||  
 Qy 958 GCGCATATCTCGCT-----CGATTATCTATCTCTCGAGGA 990  
 Db |||||  
 Qy 282 GCCACATATCTCCCTCAAGGCTTAGAGACTAATCTCTAGTGTCTATTTTCACTCTCGAGGA 223  
 Db |||||  
 Qy 991 ACCGATCTAGGAGCTCTAGTTCAGCTTTTACCTTTCAGTTTGAAGAAATATCTAGTGT 1050  
 Db |||||  
 Qy 222 ACCGATCTAGGAGGTTCAAACTCAGCTTTTACTTTCATGTTATGAGAAATCTAAACAGGTTT 163  
 Db |||||  
 Qy 1051 CTGTGATGGTGAATTAATAACTCGGAGGCGGGTTTCGGATCTACATTTTCAAGAGAGAGCA 1110  
 Db |||||  
 Qy 162 CTGTGATGGCGTTTC-----GCANAGACAAGCA 136  
 Db |||||  
 Qy 1111 TTAGAGCAAAAGAAACCCATCTCTTGGATCTTCTTCAAAATGGT 1154  
 Db |||||  
 Qy 135 TTAGAGCAAAAGAAACCCATCTCTTGGATCTTCTTCAAAATGGT 92  
 Db |||||

RESULT 12

BUS37920 751 bp mRNA linear EST 16-OCT-2002  
 LOCUS T107B12 Populus apical shoot cDNA library Populus tremula x Populus  
 DEFINITION tremuloides cDNA 5 prime, mRNA sequence.  
 ACCESSION BUS37920  
 VERSION BUS37920.1 GI:24020716  
 SOURCE EST.  
 ORGANISM Populus tremula x Populus tremuloides  
 Populus tremula x Populus tremuloides  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Malpighiales; Salicaceae; Populus.  
 REFERENCE 1 (bases 1 to 751)  
 AUTHORS Unneberg,P., Bhalerao,R.R., Jansson,S. and Sterky,F.  
 TITLE The poplar tree transcriptome: Analysis of expressed sequence tags  
 from multiple libraries  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: BHALERAO RUPALI R.  
 Umea Plant Science Center  
 Department of Plant Physiology  
 University of Umea, 901 87 Umea, Sweden  
 Tel: +46 90 786 5279  
 Fax: +46 90 786 6676  
 Email: rupali.bhalerao@plantphys.umu.se.  
 Location/Qualifiers  
 1..751  
 /organism="Populus tremula x Populus tremuloides"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:47664"  
 /tissue\_type="apical shoot"  
 /clone\_lib="Populus apical shoot cDNA library"

ORIGIN

Query Match 11.5%; Score 216.4; DB 13; Length 751;  
 Best Local Similarity 68.8%; Pred. No. 4.8e-31;  
 Matches 333; Conservative 0; Mismatches 136; Indels 15; Gaps 2;  
 Qy 1094 CATTTCAGAGGAGCATTAGAAGCAAGAAACCAATCTCTTGGATCTTCTTCAATGG 1153  
 Db |||||  
 Qy 271 CACTGCAGAAACCGGCACCTGGAAGCAAGAGAAACAAATGCTAATCTACTTCAAGTGG 330  
 Db |||||  
 Qy 1154 TGGATGATCGATATAGTTCATTGCGTAGATGAGATTATACAGGTTATATCAGGTTCCATG 1213  
 Db |||||  
 Qy 331 TTGATGAACGATATAGCAATGCTTGGATGAATAACACAGATTATTCAGCATTCATG 390  
 Db |||||  
 Qy 1214 CTGCAACCGAGTTAGATCCACAGTTACACACCGGTTTGCCTCCAAACGGTTTCCTTCT 1273  
 Db |||||  
 Qy 391 CTGCAACTGAGTTGGACCCACAGATTACACAGTTTTTCTTCTTCAACAATACTTTCT 450  
 Db |||||  
 Qy 1274 TATACAAAGACCTCGAGAGAGAGAAATCTGCAAGAGAGATAATCTCTATGGATCTGTATTG 1333  
 Db |||||  
 Qy 451 TGTACAAACGCTTAGGAGCGGATCAGCAACCAATCTAGCAATGGAGGTCATTGG 510  
 Db |||||  
 Qy 1334 AGAG-----AGCAAGACAGACTCAAGAAACCTTATGTTCCACCAGCAT 1381  
 Db |||||  
 Qy 511 ATAGTGAGACACATAGAGACAGAAAGGATCTTTTGGAACTTCATACTTACAAAAGCAAT 570  
 Db |||||  
 Qy 1382 GCCTTCTTACAGAGCTGAACGAAAGAACCATCAGATTTTGGAGACCTTCAACAGGTTTTC 1441  
 Db |||||  
 Qy 571 GGACTCTCCAGAGCTGAGAGAAATGACCAAC---TGTGGAGACCCAGAGAGGCTTAC 627  
 Db |||||  
 Qy 1442 CTGAGAAATCTGTTTGGTTCTACCGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 1501  
 Db |||||  
 Qy 628 CAGAAAGATCTGCTCAGTTCTCGGTGCATGGATGTTTTCAGAACTTTTTCACCCGTACC 687  
 Db |||||  
 Qy 1502 CGAAAGATTCGAGAGAAACATCTTCTAGCTATACGAAGTGGCTTGACAAGAGTCAGGTAT 1561  
 Db |||||  
 Qy 688 CTAAAGATTCAGAGAGCAATTTGCTTGGCGCAAAAGTGGACTTAACAGAGCCAGGTAT 747  
 Db |||||  
 Qy 1562 CAAA 1565  
 Db |||||  
 Qy 748 CAAA 751  
 Db |||||

1548 AGAAGTCAGGTATCAAACTGGTTTATAAATCGCGGGTTAGGCTATGAAGCCGATGAT 1607  
|||||  
260 GAGAGTCAGGTTTCAAACTGGTTTATAAATCGTAGAGTTGGCTATGAAGCCATGGT 319  
|||||  
1608 AGAAGAGATGATCGGAATGACAAAGAGGAGCTCAATACAG 1652  
|||||  
320 AGAAGAGATGATCTACTGGAAGATGCAAGGAACATGAACACAGACAG 364  
|||||

RESULT 14  
BU024863 672 bp mRNA linear EST 23-AUG-2002  
LOCUS  
DEFINITION QH6H21.yg.ab1 QH EFGHJ sunflower RHA280 Helianthus annuus cDNA  
clone QH6H21, mRNA sequence.  
ACCESSION BU024863  
VERSION BU024863.1 GI:22460383  
KEYWORDS  
SOURCE Helianthus annuus (common sunflower)  
ORGANISM Helianthus annuus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; campanulids; Asterales; Asteraceae; Asteroideae;  
Heliantheae; Helianthus.  
1 (bases 1 to 672)  
Kozik,A., Michelmoro,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,  
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J.,  
Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,  
Lai,Z., Church,S., Jackson,L. and Bradford,K.  
Lettuce and Sunflower ESTs from the Compositae Genome Project  
http://comgenomics.ucdavis.edu/  
Unpublished (2002)  
Contact: Alexander Kozik [R.W.Michelmoro]  
Department of Vegetable Crops, R.W.Michelmoro Lab  
University of California at Davis (UCD)  
Asmundson Hall, UCD, Davis, CA 95616, USA  
Tel: 1-(530)-742-1742  
Fax: 1-(530)-752-9659  
Email: akozik@ucdavis.org [michelmoro@vegmail.ucdavis.edu]  
Belongs to contig QH\_CA\_Contig2963, see http://cgdb.ucdavis.edu/  
for details.  
Plate: QH6 row: H column: 21.

FEATURES  
Location/Qualifiers  
1..672  
/organism="Helianthus annuus"  
/mol\_type="mRNA"  
/cultivar="RHA280"  
/db\_xref="taxon:4232"  
/clone="QH6H21"  
/lab\_host="E.coli"  
/note="lib="QH EFGHJ sunflower RHA280"  
/note="Vector: pBRCDNASfiAB; The library was constructed  
from 11 different sources of RNA from a single genotype.  
Separate cDNAs were generated using primers that  
incorporated unique 5' and 3' tags to distinguish each  
source of RNA. cDNAs were then pooled, size-fractionated,  
directionally cloned into a custom medium-copy vector and  
transformations made with four size classes to minimize  
size bias. Details of each source of RNA and library  
construction can be obtained at http://cgdb.ucdavis.edu/  
TAG\_SEQ=GTAGTCGGG"

ORIGIN  
Query Match 7.8%; Score 147.4; DB 13; Length 672;  
Best Local Similarity 69.8%; Pred. No. 6e-18;  
Matches 199; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 1368 GTTCCACGAGATTGCTTCTTTCAGCAGCTGAAACGAAAGAACCATCAGATTGGAGACC 1427  
Db 80 GCTCCGCGACACGCGGCTTTACAGCAATAGGAATGATCCACACAAATGCTTTGGAGACC 139  
QY 1428 TCAACGAGGTTTCCCTGAGAAATCTCTTTCGGTTTCTACGGAATGGATCTTCCAAAACCTT 1487  
Db 140 ACAAGAGGATTCCCGAACCGCTCTCTTTCGGTTCTTTCGTTGCTGCTTTTGAACACTT 199  
QY 1488 CTTTCACCTTACCCGAAAGATTCCGAGAAACATCTTCTAGCTATACGAGTGGCTTGAC 1547  
Db 200 TCTCCATCTTATCCAAAGATTCCGATAAACAATATGCTGGGAAAACAACTGGACTCAC 259

1548 AGAAGTCAGGTATCAAACTGGTTTATAAATCGCGGGTTAGGCTATGAAGCCGATGAT 1607  
|||||  
260 GAGAGTCAGGTTTCAAACTGGTTTATAAATCGTAGAGTTGGCTATGAAGCCATGGT 319  
|||||  
1608 AGAAGAGATGATCGGAATGACAAAGAGGAGCTCAATACAG 1652  
|||||  
320 AGAAGAGATGATCTACTGGAAGATGCAAGGAACATGAACACAGACAG 364  
|||||

RESULT 14  
BU026692 698 bp mRNA linear EST 23-AUG-2002  
LOCUS  
DEFINITION QH6H21.yg.ab1 QH EFGHJ sunflower RHA280 Helianthus annuus cDNA  
clone QH6H21, mRNA sequence.  
ACCESSION BU026692  
VERSION BU026692.1 GI:22462212  
KEYWORDS  
SOURCE Helianthus annuus (common sunflower)  
ORGANISM Helianthus annuus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; campanulids; Asterales; Asteraceae; Asteroideae;  
Heliantheae; Helianthus.  
1 (bases 1 to 698)  
Kozik,A., Michelmoro,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,  
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J.,  
Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,  
Lai,Z., Church,S., Jackson,L. and Bradford,K.  
Lettuce and Sunflower ESTs from the Compositae Genome Project  
http://comgenomics.ucdavis.edu/  
Unpublished (2002)  
Contact: Alexander Kozik [R.W.Michelmoro]  
Department of Vegetable Crops, R.W.Michelmoro Lab  
University of California at Davis (UCD)  
Asmundson Hall, UCD, Davis, CA 95616, USA  
Tel: 1-(530)-742-1742  
Fax: 1-(530)-752-9659  
Email: akozik@ucdavis.org [michelmoro@vegmail.ucdavis.edu]  
Belongs to contig QH\_CA\_Contig2963, see http://cgdb.ucdavis.edu/  
for details.  
Plate: QH6 row: L column: 13.

FEATURES  
Location/Qualifiers  
1..698  
/organism="Helianthus annuus"  
/mol\_type="mRNA"  
/cultivar="RHA280"  
/db\_xref="taxon:4232"  
/clone="QH6H21"  
/lab\_host="E.coli"  
/note="lib="QH EFGHJ sunflower RHA280"  
/note="Vector: pBRCDNASfiAB; The library was constructed  
from 11 different sources of RNA from a single genotype.  
Separate cDNAs were generated using primers that  
incorporated unique 5' and 3' tags to distinguish each  
source of RNA. cDNAs were then pooled, size-fractionated,  
directionally cloned into a custom medium-copy vector and  
transformations made with four size classes to minimize  
size bias. Details of each source of RNA and library  
construction can be obtained at http://cgdb.ucdavis.edu/  
TAG\_SEQ=Not found"

ORIGIN  
Query Match 7.8%; Score 147.4; DB 13; Length 698;  
Best Local Similarity 69.8%; Pred. No. 5.9e-18;  
Matches 199; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 1368 GTTCCACGAGATTGCTTCTTTCAGCAGCTGAAACGAAAGAACCATCAGATTGGAGACC 1427  
Db 108 GCTCCGCGACACGCGGCTTTACAGCAATAGGAATGATCCACACAAATGCTTTGGAGACC 167  
QY 1428 TCAACGAGGTTTCCCTGAGAAATCTCTTTCGGTTTCTACGGAATGGATGTTCAAAACCTT 1487  
Db 168 ACAAGAGGATTCCGGAACGCTCTGTTTCGTTCTTTCGTTGCTGCTTTTGAACACTT 227

QY 1488 CCTTACCCCTTACCGAAGATTCGGAGAAACATCTTTAGCTATATACGAAGTGGCTTGAC 1547  
 Db 228 TCTCCATCCTTATCCAAAGATTTCGATAAAACATATGCTGGCAAAACAACCTGGACTCAC 287  
 QY 1548 AAGAAGTCAAGGATCAAACTGTTTATAAATCGCGGGTAGGCTATGGAAGCCGATGAT 1607  
 Db 288 GAGAAGTCAAGGTTCAAACTGTTTATAAATCGTAGAGTTCGGCTATGGAAGCCCAATGGT 347  
 QY 1608 AGAAGAGATGTATCGGGAATGAACAAGAGGAGCTCAATAACAG 1652  
 Db 348 AGAAGAGATGTACTTGGAGAATGCAAGGAACATGAACAGAACAG 392

## RESULT 15

AL946255  
 LOCUS 328 bp DNA linear GSS 24-OCT-2002  
 DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-295B12-015445,  
 genomic survey sequence.  
 ACCESSION AL946255  
 VERSION AL946255.1 GI:24402877  
 KEYWORDS GSS.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 REFERENCE 1 Strizhov,N., Li,Y., Rosso,M., Viehoveer,P., Dekker,K., Saedler,H.  
 and Weisshaar,B.  
 A pipeline for automated high-throughput generation of FSTs  
 (flanking sequence tags) from Arabidopsis thaliana T-DNA  
 transformed lines  
 Unpublished  
 JOURNAL  
 REFERENCE 2  
 AUTHORS Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.  
 TITLE A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)  
 for flanking sequence tag based reverse genetics  
 Unpublished  
 JOURNAL  
 REFERENCE 3 (bases 1 to 328)  
 AUTHORS Li,Y., Strizhov,N., Rosso,M. and Weisshaar,B.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-OCT-2002) Weisshaar B., Max-Planck-Institut fuer  
 Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany  
 This sequence is recovered from the left border of the T-DNA. It  
 indicates an insertion within the locus defined by clone F26p21.  
 The sequences are generated at the MPI for Plant Breeding Research  
 in the context of the GABI-Kat project. GABI-Kat is part of the  
 German Plant Genomics program designated 'GABI'. Information on  
 line availability can be found at:  
 http://www.mpiz-koeln.mpg.de/GABI-Kat/  
 Location/Qualifiers

## FEATURES

source  
 1..328  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia 0"  
 /db\_xref="taxon:3702"  
 /clone\_lib="Arabidopsis thaliana T-DNA insertion lines"  
 /note="PCR was performed on DNA from Arabidopsis thaliana  
 plants (T1) which were transformed with the T-DNA from  
 vector PAC161. The lines contain one or more T-DNA  
 insertions. The DNA fragment(s) resulting from the PCR  
 were directly sequenced to determine the genomic sequence  
 flanking the insertion. Sequences displaying significant  
 similarity to the A. thaliana nuclear genome sequence were  
 processed for submission. T-DNA derived sequences were  
 removed"

## ORIGIN

Query Match 7.7%; Score 146; DB 29; Length 328;  
 Best Local Similarity 98.8%; Pred. No. 1.4e-17;  
 Matches 168; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 119 CTGCGAATGAAAAGTACACTGTTCTTTTCACGGAGAAAGAGATAAAATAAGCATTATCTTC 178  
 Db 25 CTGCGAATGAAAAGT-CACGTGTTCTTTTCACGGAGAAAGAGATAAAATAAGCATTATCTTC 83  
 QY 179 TTCTTTCAAGTTTTTAACACACACATTTTGGAAATTTTGATGTAAAAATCTCTTTGGAACGTT 238  
 Db 84 TTCTT-AGTTTTTAACACACATTTTGGAAATTTTGATGTAAAAATCTCTTTGGAACGTT 142  
 QY 239 GTGTTGTCGTAATCTTCCAAAGGTTCTATCAGAAGAAGAGGATAAAG 288  
 Db 143 GTGTTGTCGTAATCTTCCAAAGGTTCTATCAGAAGAAGAGGATAAAG 192

Search completed: September 2, 2004, 09:48:35  
 Job time : 4957 secs